

**UNIVERSIDADE FEDERAL DO PARANÁ**

**ADEMIR LUIZ DO PRADO**

**ESTUDO DA APLICAÇÃO DE REDES NEURONAIS ARTIFICIAIS  
PARA APOIO À DECISÃO NA LIBERAÇÃO DO  
PERFIL LIPÍDICO E DE GLICEMIA EM JEJUM**

**CURITIBA**

**2011**

**ADEMIR LUIZ DO PRADO**

**ESTUDO DA APLICAÇÃO DE REDES NEURONAIS ARTIFICIAIS  
PARA APOIO À DECISÃO NA LIBERAÇÃO DO  
PERFIL LIPÍDICO E DE GLICEMIA EM JEJUM**

Dissertação apresentada ao Programa de Pós-Graduação em Bioinformática, Área de Concentração em Bioinformática, Setor de Educação Profissional e Tecnológica, Universidade Federal do Paraná, como parte dos requisitos para obtenção do título de Mestre em Bioinformática.

Orientador: Profa. Dra. Jeroniza Nunes Marchaukoski  
Co-orientador: Prof. Dr. Geraldo Picheth

**CURITIBA**

**2011**

Prado, Ademir Luiz do

Estudo da aplicação de redes neuronais artificiais para apoio à  
decisão na liberação do perfil lipídico e de glicemia em jejum /  
Ademir Luiz do Prado – Curitiba, 2011.

99 f. il. (algumas color.); 30 cm.

Orientadora: Professora Dra. Jeroniza Nunes Marchaukoski

Co-orientador: Professor Dr. Geraldo Picheth

Dissertação (mestrado) – Programa de Pós-Graduação em  
Bioinformática, Setor de Educação Profissional e Tecnológica,  
Universidade Federal do Paraná.

Inclui bibliografia

1. Redes neurais. 2. Perfil lipídico. 3. Glicemia em jejum.  
I. Marchaukoski, Jeroniza Nunes. II. Picheth, Geraldo. III.  
Universidade Federal do Paraná. IV. Título.

CDD 006.32

## TERMO DE APROVAÇÃO

**ADEMIR LUIZ DO PRADO**

Estudo da aplicação de redes neuronais artificiais para apoio à decisão na liberação do perfil lipídico e de glicemia em jejum

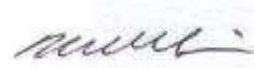
Dissertação aprovada como requisito parcial para obtenção do grau de Mestre em Bioinformática, pelo Programa de Pós-graduação em Bioinformática, Setor de Educação Profissional e Tecnológica, da Universidade Federal do Paraná, pela seguinte banca examinadora:

Orientadora:

  
Prof.ª Dr.ª Jeroniza Nunes Marchaukoski

Coorientador:

  
Prof. Dr. Geraldo Picheth

  
Prof. Dr. Roberto Tadeu Raittz  
Universidade Federal do Paraná

  
Prof.ª Dr.ª Marileia Scartezini  
Universidade Federal do Paraná

Curitiba, 21 de fevereiro de 2011

## DEDICATÓRIA

*Dedico este trabalho à minha esposa Leocádia pela atenção, carinho, cuidado e compreensão e as minhas filhas Amanda e Larissa, alvos de minha dedicação e empenho.*

## AGRADECIMENTOS

Primeiramente ao Senhor Jesus em quem eu creio e me reconciliou eternamente com Deus. Ele me fortaleceu todas as vezes que O invoquei. Aos irmãos da Igreja em Curitiba que oraram por mim.

A minha esposa Leocádia pelo apoio, carinho e cuidado comigo. As minhas filhas Amanda e Larissa que às vezes não entendiam a minha ausência, mas aceitavam a necessidade que eu tinha de estudar.

A minha querida mãe, Leonilda, exemplo de abnegação e dedicação aos filhos. Ao meu pai Pedro (*in memoriam*). Aos meus irmãos Claudinei e Claudiane. Aos meus tios, Dorival e Lúcia que me receberam e me acolheram nesta cidade.

A professora, orientadora e doutora Jeroniza Nunes Marchaukoski, pelo direcionamento e esclarecimento prestados. Ao professor, orientador e doutor Geraldo Picheth, que me inspirou no tema abordado, pelo encorajamento, amizade, suporte e sabedoria na condução do trabalho. A professora e doutora Mariléia Scartezini pelas ricas sugestões sobre os exames estudados. Ao professor e doutor Roberto Tadeu Raittz pelo suporte em redes neuronais artificiais.

Ao professor e doutor José Simão de Paula Pinto pela amizade e incentivo bem como a carta de recomendação. Ao professor e doutor Admar Moraes de Souza pela carta de recomendação. Ao professor e mestre Everaldo Leonel de Oliveira (*in memoriam*) pela carta de recomendação e pelo incentivo na carreira docente.

Aos professores do programa de Pós-Graduação em Bioinformática pelo conhecimento compartilhado. As secretárias Mariana Marques Pereira, Léa Regina Welter e Suzana de Azevedo Gobetti pela disposição e atenção de sempre.

Aos meus colegas de mestrado Daniela, Dieval, Eduardo, Guilherme, Leandro, Luiz Fernando, Michelly, Rosa, Rodrigo, Terumi, Waldir pela amizade e companheirismo, em especial ao colega Waldemar Volanski por compartilhar sua experiência sobre o tema abordado.

Ao Hospital de Clínicas da Universidade Federal do Paraná pela oportunidade de realizar o trabalho baseado nos resultados dos exames de glicemia em jejum e perfil lipídico. Aos colegas do laboratório do Hospital de Clínicas, os bioquímicos Carlos e Kátia pela participação como avaliadores dos exames laboratoriais clínicos.

A todas as pessoas que torceram por mim e de alguma forma me ajudaram na realização deste trabalho.

*“Bem-aventurado o homem que acha sabedoria, e o homem que adquire conhecimento; porque é melhor a sua mercadoria do que artigos de prata, e maior o seu lucro que o ouro mais fino”.*

Provérbios 3:13-14

## RESUMO

As determinações do perfil lipídico (colesterol total, HDL-colesterol, LDL-colesterol, triglicérides) e da glicemia em jejum são ensaios de grande demanda nos laboratórios clínicos. A liberação destes resultados por profissionais consome tempo e atenção. O estudo se propõe avaliar a aplicação das redes neurais, *Multilayer Perceptron* (MLP) e *Free Associative Neurons* (FAN), como ferramentas de inteligência artificial para colaborar na liberação dos resultados, em processo designado “segunda opinião”. O projeto tem a aprovação do Comitê de Ética em Pesquisa com Seres Humanos do HC-UFPR (CAE: 0253.0.208.000-10). Uma amostra contendo 60.006 registros obtidos do banco de dados do HC-UFPR foi analisada. A idade média dos pacientes foi cerca de 47 anos (amplitude de variação de 2 a 99 anos), com predomínio de mulheres (~65%). Esta amostra foi classificada em “liberado” quando todos os valores dentro dos critérios estabelecidos de normalidade e “retido” quando qualquer analito estudado se mostrou fora da referência. Esta classificação resultou em 62% da amostra classificada no grupo “retido”. Quando as redes neurais foram testadas com arquivos completos (n=30.003) a rede FAN apresentou divergência cerca de 6 vezes superior à rede MLP (7,6% vs. 1,2%) embora ambas tenham um desempenho satisfatório em acurácia (>90%). Foram treinadas e testadas as redes FAN e MLP com arquivos incompletos, caracterizados pela ausência de algum dos parâmetros em estudo com diferentes tamanhos de arquivos (30.536, 65.536 e 120.000 registros). Nesta condição que mimetiza os resultados liberados pelo laboratório clínico, a rede neuronal MLP apresentou desempenho superior à rede FAN. O estudo permitiu concluir que: (1) a rede neuronal FAN perde desempenho com arquivos incompletos, (2) a rede neuronal MLP apresentou desempenho superior à rede FAN quando estudada com arquivos completos ou incompletos, (3) o tamanho amostral utilizado para treinamento e teste não afetaram o desempenho da rede neuronal MLP, enquanto que a rede FAN é afetada por perda de sensibilidade, (4) resultados divergentes da rede neuronal MLP avaliados por especialistas humanos evidenciaram que os ensaios com valores alterados foram o principal elemento de inconsistência. Em síntese, a rede neuronal MLP é recomendada para outros estudos com desenho amostral semelhante e apresenta potencial para aplicação no laboratório clínico como suporte a decisão na liberação de resultados.

Palavras-Chaves: Perfil lipídico. Glicemia em jejum. Redes neurais. MLP. FAN.



## ABSTRACT

The lipid profile (total cholesterol, HDL-cholesterol, LDL-cholesterol, triglycerides) and fasting blood glucose are tests of high throughput in clinical laboratories. In the process of liberate these results professionals needs to takes time and attention. The study aims to evaluate the application of neural networks, Multilayer Perceptron (MLP) and Free Associative Neurons (FAN), as artificial intelligence tools to assist in the release of the results, a process called "second opinion". The project was approved by the Ethic Committee in Human Research of the HC-UFPR (CAE: 0253.0.208.000-10). A sample containing 60,006 registers obtained from HC-UFPR Database was analyzed. The mean age of patients was about 47 years (range: 2-99 years) with predominance of women (~65%). This sample was classified as "released", when all values were within the established criteria of normality and "retained" when any studied analyte showed values outside the reference of normality. This classification resulted in 62% of the sample classified as "retained". When neural networks were tested with complete files (n=30,003) FAN network disagreement was about 6 times higher compared with the MLP network (7.6% vs. 1.2%) although both have a satisfactory performance in accuracy (> 90%). The FAN and MLP networks were also trained and tested with incomplete files, characterized by the absence of any of the parameters under study with different file sizes (30,536, 65,536 and 120,000 registers). In this condition that mimics the results released by a clinical laboratory, the MLP neural network showed a superior performance compared to the FAN network. The study concluded that: (1) the FAN neural network loses performance with incomplete files, (2) the performance of MLP neural network was superior to the FAN network when tested with complete or incomplete files, (3) the sample size used to training and testing did not affect the performance of MLP neural network, while FAN network that is affected by loss of sensitivity, (4) divergent results of the MLP neural network evaluated by human experts showed that the tests with high values were the main element of inconsistency. In summary, the MLP neural network is recommended for other studies with similar sample design and presents a potential for application in the clinical laboratory as a decision support system.

Key Words: Lipid profile. Fasting glucose. Neural networks. MLP. FAN.

## LISTA DE FIGURAS

FIGURA 1 - NÚMERO DE ÓBITOS POR DOENÇA.....	22
FIGURA 2 - HOSPITAL DE CLÍNICAS DA UFPR.....	23
FIGURA 3 - ANALISADOR AUTOMÁTICO ARCHITECT ci8200 .....	26
FIGURA 4 - EXEMPLO DE REQUISIÇÃO INFORMATIZADA DE EXAMES LABORATORIAIS NO SIH .....	29
FIGURA 5 - EXEMPLO DE REGISTRO DE RESULTADO INFORMATIZADO DE EXAMES LABORATORIAIS NO SIH .....	30
FIGURA 6 - EVOLUÇÃO DO PROCESSO ATEROSCLERÓTICO.....	32
FIGURA 7 - EXEMPLO DE TUBO DE COLETA DE SANGUE COM ETIQUETA DE CÓDIGO DE BARRAS PARA INTERFACEAMENTO.....	34
FIGURA 8 - FLUXOGRAMA DAS ETAPAS DO INTERFACEAMENTO.....	36
FIGURA 9 - EXEMPLO DE INTERFACE COM <i>LOGIN</i> DE USUÁRIO NO SISTEMA DE INTERFACEAMENTO.....	37
FIGURA 10 - EXEMPLO DE INTERFACE DO PROCESSO DE IMPORTAÇÃO E EXPORTAÇÃO DE EXAMES DO SISTEMA DE INTERFACEAMENTO.....	38
FIGURA 11 - REPRESENTAÇÃO SIMPLIFICADA DE UM NEURÔNIO BIOLÓGICO.....	41
FIGURA 12 - ESTRUTURA ESQUEMÁTICA DO MODELO MATEMÁTICO DE UM NEURÔNIO.....	42
FIGURA 13 - REPRESENTAÇÃO SIMPLIFICADA DE UMA REDE NEURONAL ARTIFICIAL.....	43
FIGURA 14 - CLASSIFICAÇÃO DOS ARQUIVOS DE DADOS UTILIZADOS PARA ESTUDOS COM REDES NEURONAIS.....	52
FIGURA 15 - CLASSIFICAÇÃO DOS EXAMES PARA REDE NEURONAL.....	54
FIGURA 16 - AMBIENTE DE DESENVOLVIMENTO DO <i>SOFTWARE</i> MATLAB® .....	62
FIGURA 17 - DEMONSTRATIVO DA TELA DO <i>SOFTWARE</i> EASYFAN.....	63
FIGURA 18 - EXEMPLO DO <i>SOFTWARE</i> OPENEPI.....	66
FIGURA 19 - HISTOGRAMA DA VARIÁVEL IDADE NA AMOSTRA CLASSIFICADA NOS GRUPOS EM ESTUDO. ....	70

## LISTA DE TABELAS

TABELA 1 - ESTRUTURA GERAL DO HOSPITAL DE CLÍNICAS DA UFPR .....	24
TABELA 2 - VALORES DE REFERÊNCIA PARA O PERFIL LIPÍDICO.....	32
TABELA 3 - EFEITO DO INTERFACEAMENTO EM LABORATÓRIOS DE ANÁLISES CLÍNICAS.....	35
TABELA 4 - ESTUDOS COM REDES NEURONAIS PARA IDENTIFICAÇÃO DE PADRÕES ASSOCIADOS AO DIGNÓSTICO CLÍNICO.....	48
TABELA 5 - CRITÉRIOS PARA A CLASSIFICAÇÃO DA AMOSTRA .....	53
TABELA 6 - EXEMPLO DA DISTRIBUIÇÃO DOS DADOS USADOS PARA TREINAMENTO DA REDE PARA ARQUIVO COMPLETO (C-1) .....	55
TABELA 7 - EXEMPLO DA DISTRIBUIÇÃO DOS DADOS USADOS PARA TREINAMENTO DA REDE PARA ARQUIVO INCOMPLETO (I-1) .....	56
TABELA 8 - ESTRUTURA DA MATRIZ DE DIVERGÊNCIA UTILIZADA NO PROJETO .....	65
TABELA 9 - ANÁLISE DE NORMALIDADE PARA AS VARIÁVEIS EM ESTUDO.....	67
TABELA 10 - CARACTERÍSTICAS GERAIS DO BANCO DE DADOS MYSQL UTILIZADO NO ESTUDO. ....	68
TABELA 11 - CARACTERÍSTICAS DA AMOSTRA EM ESTUDO.....	69
TABELA 12 - ANÁLISE DE CORRELAÇÃO LINEAR (PEARSON) COM OS PARÂMETROS EM ESTUDO.....	71
TABELA 13 - MATRIZ DE DIVERGÊNCIA PARA AS REDES NEURONAIS MLP E FAN.....	72
TABELA 14 - COMPARAÇÃO DAS REDES NEURONAIS COM ARQUIVOS COMPLETOS .....	72
TABELA 15 - MATRIZ DE DIVERGÊNCIA PARA A REDE FAN COM ARQUIVOS INCOMPLETOS.....	73
TABELA 16 - MATRIZ DE DIVERGÊNCIA PARA A REDE NEURONAL MLP COM ARQUIVOS INCOMPLETOS .....	74
TABELA 17 - COMPARAÇÃO ENTRE AS REDES MLP E FAN COM ARQUIVOS INCOMPLETOS.....	75
TABELA 18 - COMPARAÇÃO ENTRE ANALISTAS HUMANOS COM OS RESULTADOS DIVERGENTES DA REDE MLP.....	75

## LISTA DE ABREVIATURAS

AIX	- <i>Advanced Interactive eXecutive</i>
AVC	- Acidente Vascular Cerebral
CAD	- <i>Computer-Aided Diagnosis</i>
CSP	- <i>Cross Structured Program</i>
CT	- Colesterol Total
DIC	- Doença Isquêmica do Coração
FAN	- <i>Free Associative Neurons</i>
FNN	- <i>Fuzzy Neural Network</i>
FNS	- <i>Fuzzy Neural System</i>
FSF	- <i>Free Software Foundation</i>
GPL	- <i>General Public License</i>
HC-UFPR	- Hospital de Clínicas da UFPR
HD	- <i>Hard Disk</i>
HDL	- <i>High Density Lipoprotein</i>
IA	- Inteligência Artificial
LDL	- <i>Low Density Lipoprotein</i>
LM	- <i>Levenberg-Marquardt</i>
MATLAB	- <i>Matrix Laboratory</i>
MLP	- <i>Multilayer Perceptron</i>
NHLBI	- <i>National Heart Lung and Blood Institute</i>
OMS	- Organização Mundial da Saúde
QMF	- <i>Query Management Facility</i>
RNA	- Rede Neuronal Artificial
RP	- Reconhecimento de Padrões
SAC	- Serviço de Análises Clínicas
SAD	- Sistema de Apoio a Decisão
SBC	- Sistemas Baseados em Conhecimento
SGBD	- Sistema Gerenciador de Banco de Dados
SI	- Sistema de Informação
SIH	- Sistema de Informações Hospitalares
SO	- Sistema Operacional
SQL	- <i>Structured Query Language</i>
TG	- Triglicérides
TOTG	- Teste Oral de Tolerância à Glicose
UAD	- Unidade de Apoio Diagnóstico
UI	- Unidade de Informação
VLDL	- <i>Very Low Density Lipoprotein</i>

## SUMÁRIO

<b>1 INTRODUÇÃO</b>	13
1.1 CONSIDERAÇÕES INICIAIS	13
1.2 JUSTIFICATIVA DO TRABALHO	16
1.2.1 Metas para a Aplicação das Redes Neurais neste Projeto	17
<b>2 OBJETIVOS</b>	19
2.1 OBJETIVO GERAL	19
2.2 OBJETIVOS ESPECÍFICOS	19
<b>3 REVISÃO DE LITERATURA</b>	20
3.1 O ESTUDO DE FRAMINGHAM	20
3.2 HOSPITAL DE CLÍNICAS DA UNIVERSIDADE FEDERAL DO PARANÁ	23
3.2.1 Unidade de Apoio Diagnóstico	25
3.2.2 Unidade de Informação	26
3.3 PERFIL LIPÍDICO E GLICEMIA EM JEJUM	31
3.3.1 Glicemia em Jejum	31
3.3.2 Perfil Lipídico	31
3.4 INFORMÁTICA E O LABORATÓRIO DE ANÁLISES CLÍNICAS	34
3.4.1 Interfaceamento no Hospital de Clínicas	36
3.5 INTELIGÊNCIA ARTIFICIAL	39
3.5.1 Reconhecimento de Padrões	39
3.5.2 Redes Neurais Artificiais	40
3.5.2.1 <i>Multilayer Perceptron</i>	44
3.5.2.2 <i>Neuro-Fuzzy</i>	46
3.5.3 Integração em Inteligência Computacional	46
3.6 TRABALHOS APROXIMADOS COM REDES NEURONAIS ARTIFICIAIS	47
<b>4 MATERIAIS E MÉTODOS</b>	50
4.1 AMOSTRA	50
4.2 CLASSIFICAÇÃO DA AMOSTRA	53
4.3 DOSAGENS LABORATORIAIS	56
4.4 EQUIPAMENTOS E SOFTWARES UTILIZADOS	57
4.4.1 <i>Hardware</i>	57
4.4.2 <i>Softwares</i>	57
4.4.2.1 DB2®	57

4.4.2.2 SQL .....	58
4.4.2.3 EXCEL® .....	58
4.4.2.4 WAMPSEVER .....	59
4.4.2.5 MYSQL .....	59
4.4.2.6 STATISTICA® .....	60
4.4.2.7 OPENEPI .....	61
4.4.2.8 MATLAB® .....	61
4.4.2.9 EASYFAN.....	63
4.4.3 Treinamento e Teste das Redes Neurais .....	64
4.5 DIVERGÊNCIAS DA REDE MLP ANALISADAS POR PROFISSIONAIS ESPECIALISTAS.....	64
4.6 ANÁLISES ESTATÍSTICAS.....	65
<b>5 RESULTADOS</b> .....	68
5.1 AMOSTRA.....	68
5.1.1 Características do Banco de Dados MYSQL.....	68
5.1.2 Características da Amostra em Estudo .....	69
5.2 ANÁLISE DAS REDES NEURONAS TREINADAS .....	71
5.2.1 Análise da Rede Neuronal FAN com Registros Incompletos .....	73
5.2.2 Análise da Rede Neuronal MLP com Registros Incompletos .....	73
5.3 RESULTADOS ANALISADOS POR PROFISSIONAIS ESPECIALISTAS .....	75
<b>6 DISCUSSÃO</b> .....	76
6.1 AMOSTRA.....	77
6.2 ANÁLISE DAS REDES NEURONAS .....	79
6.2.1 Análise das Redes Neurais com Arquivos Completos.....	79
6.2.2 Análise das Redes Neurais com Arquivos Incompletos .....	80
6.3 PERSPECTIVAS FUTURAS .....	82
<b>7 CONCLUSÕES</b> .....	83
<b>REFERÊNCIAS</b> .....	84
<b>APÊNDICES</b> .....	89
<b>ANEXO</b> .....	99

# 1 INTRODUÇÃO

Os usuários de informática possuem avançadas tecnologias que permitem aprimorar seus trabalhos, aumentar a qualidade e como consequência aumentar o estímulo pela possibilidade do acesso às novas tecnologias tais como equipamentos laboratoriais e *softwares* que dão apoio para o controle de exames.

Os Sistemas de Informações (SI) são utilizados em todas as médias e grandes empresas para gerenciar informações das mais variadas áreas desde a bancária até laboratório de análises clínicas.

## 1.1 CONSIDERAÇÕES INICIAIS

A população humana, em algum momento de suas vidas, realiza exames laboratoriais para saber os níveis de lipídeos ou glicose no organismo. A preocupação básica das pessoas é saber se o resultado está alterado conforme valores de referência para os referidos exames. Obviamente que para os médicos o interesse em saber o resultado dos exames está acima do valor em si, a preocupação é como diagnosticar possíveis doenças cardiovasculares e outras relacionadas aos exames de glicemia em jejum e perfil lipídico bem como tratá-las preventivamente.

Indivíduos que possuem valores de glicemia em jejum acima de 125 mg/dL são considerados diabéticos. De acordo com a Associação Americana de Diabetes, para ser classificado como diabético, os valores de diagnóstico são: glicemia em jejum acima de 125 mg/dL ou superior a 200 mg/dL e acompanhado de sintomas como volume urinário aumentado, sede exagerada, fome devoradora, emagrecimento inexplicável ou ainda superior a 200 mg/dL, 2 horas após ingestão de 75 g de glicose (ADA, 2010). Assim os diabéticos devem ser avaliados como doentes de alto risco cardiovascular e devem também ser avaliados relativamente à pressão arterial e aos valores dos lipídeos no sangue (colesterol, suas frações, e triglicérides). O diabetes, hipertensão arterial e dislipidemia são fatores de risco para as doenças cardiovasculares que são: Infarto Agudo do Miocárdio (IAM) e o Acidente Vascular Cerebral (AVC).

Vários trabalhos publicados apontaram as possíveis interferências nos resultados do perfil lipídico (NCEP-ATP III, 2002). Alguns cuidados devem ser observados no preparo do paciente, antecedendo à coleta de amostra, como por exemplo, o jejum prévio de 12 a 14 horas. Os valores de referência para o perfil lipídico na população brasileira encontram-se nas Diretrizes Brasileiras, publicada por especialistas da Sociedade Brasileira de Cardiologia, com a finalidade de prevenir as doenças cardiovasculares (SBC, 2007).

O laboratório de análises clínicas deixou de ser um manipulador de reagentes e amostras biológicas para incorporar novas tecnologias com processos pré-analíticos, analíticos e pós-analíticos. Tais processos são controlados através de equipamentos automáticos de alto desempenho e insumos redesenhados com técnicas de alta complexidade, obtendo-se resultados rápidos, precisos e com exatidão.

Os laboratórios de análises clínicas utilizam intensamente os sistemas informatizados. Buscam neles, além do aperfeiçoamento constante, soluções e adaptações capazes de atender às novas necessidades, auxiliando no entendimento dos resultados. Outros ganhos esperados são os relacionados com a diminuição de custos e de prazo terapêutico, maior rapidez do diagnóstico e segurança de todos os envolvidos nas ações de saúde. Os programas devem necessariamente ter a capacidade de oferecer melhorias, proporcionando o bem estar dos pacientes e dos profissionais envolvidos.

Segundo a Organização Mundial da Saúde, 17,5 milhões foram a óbito por motivo de doenças cardiovasculares e um milhão por conta do diabetes (OMS, 2005). O colesterol e suas frações são componentes lipídicos de maior importância na evolução do quadro de Aterogênese. Estudos mostram relação entre o aumento da incidência de Doença Isquêmica do Coração (DIC) e o aumento da taxa de colesterol sanguíneo, determinando então os grupos de risco. Os dados estatísticos não podem ser usados para fins terapêuticos, mas sim o nível de colesterol sérico no sangue é que pode definir prognóstico (PESCE e KAPLAN, 1987).

O Hospital de Clínicas da Universidade Federal do Paraná (HC-UFPR) possui mais de dois milhões de pacientes cadastrados e atende em média 45 mil pacientes por mês. O HC realiza através da Unidade de Apoio a Diagnóstico (UAD) uma média de 105 mil exames laboratoriais clínicos por mês (HC, 2010).



Diante dos números constatados observa-se que o laboratório de análises clínicas é uma rica fonte de dados. Esses dados interessam ao meio científico, uma vez que a sua função está relacionada com os sinais do estado da saúde humana. Entendendo a atividade que realiza o laboratório de análises clínicas e a sua importância para as modernas ações de saúde, o profissional especializado desta área tem interesse em buscar ferramentas que venham melhorar o seu desempenho não somente para o ato de fazer o exame, mas para o conjunto de variáveis que envolvem o auxílio ao diagnóstico médico. A realização dos exames demanda conhecimento, treinamento, materiais e equipamentos de última geração bem como estratégias capazes de sustentar os processos oferecendo ao paciente e ao médico resultados consistentes para avaliação diagnóstica.

O foco do profissional de análises clínicas é redirecionado para três pontos importantes da rotina de produção de resultados de exames. O primeiro, para a vigilância metodológica severa, sistemática e ininterrupta; o segundo, para o conhecimento dos mecanismos de doença, considerando a capacidade de resposta terapêutica do paciente e; o terceiro, para a disponibilidade de inúmeras outras estratégias de auxílio diagnóstico, além do laboratório clínico. Tudo isso proporciona uma massa extraordinária de conhecimento acumulado.

Os procedimentos utilizados para liberar os resultados de exames laboratoriais apresentam pontos que desfavorecem a perfeição dos resultados, como por exemplo, a validação manual realizada por profissionais, baseada em método empírico, culminando na necessidade de rever e discutir as ações da rotina das análises clínicas, visando procedimentos cada vez mais seguros que forneçam resultados fidedignos (MATSUMOTO, UEDA e KAWAJI, 2002).

Segundo Medeiros (2006), para os processos de decisão em ambientes altamente dinâmicos, já não é suficiente levar em conta as informações fornecidas pelos Sistemas de Informação. A partir do trinômio dado-informação-conhecimento, os Sistemas de Informação devem evoluir e se transformar em Sistemas Baseados em Conhecimento (SBC). A Inteligência Artificial (IA) fornece ferramentas apropriadas para suporte a sistemas inteligentes, com condições de inferir hipóteses, fornecer meios para o aprendizado da máquina e aquisição de conhecimento.

A motivação em abordar o estudo dos exames do perfil lipídico e glicemia em jejum é submetê-los com recursos de IA baseados em rede neuronal com a capacidade de liberar os resultados de exames minimizando a interferência humana.

Tal motivação é também liberar mais tempo ao profissional de laboratório para realizar pesquisa, minimizando o trabalho repetitivo de liberação de exames.

A aplicação de redes neurais para auxílio à liberação de resultados é também conhecida como “segunda opinião”, original em sua proposta para os exames de perfil lipídico e glicemia em jejum, tem como fundamento, procurar mimetizar na resposta eletrônica a decisão dos profissionais do laboratório clínico.

A realidade em que se insere este trabalho envolve os setores do HC-UFPR; que atuam em colaboração para melhor atender pacientes e médicos. Os dois principais setores são:

- Unidade de Informação (UI);
- Unidade de Apoio Diagnóstico (UAD).

Entre as áreas envolvidas estão à medicina, a informática, a bioquímica, a estatística, e outros, sendo o principal foco de todas: a área médica.

Este trabalho apresenta uma proposta da viabilidade do uso de uma rede neuronal do tipo *Multilayer Perceptron* (MLP) para liberar resultados de exames do perfil lipídico e glicemia em jejum. Também se deseja treinar, testar e comparar os dados na rede neuronal *Free Associative Neurons* (FAN). Pretende-se submeter os dados classificados no *software* desenvolvido na linguagem de programação Java chamado EasyFan produzido em 2006 (KUSTER *et al.*, 2006) e comparar com a rede neuronal MLP executada no *software* MATLAB na sua versão R2010a.

## 1.2 JUSTIFICATIVA DO TRABALHO

Existe uma demanda crescente de realização de exames laboratoriais. Cresce a população humana e aumenta a quantidade de exames que as pessoas necessitam para diagnosticar possíveis doenças. São necessários cada vez mais, Sistemas de Informações que melhorem o fluxo do processo de trabalho interno do laboratório e o fluxo dos exames nas seções até a liberação dos resultados dos mesmos.

O estudo da aplicação de redes neurais artificiais no apoio à liberação de resultados do perfil lipídico e glicemia em jejum possibilitará o desenvolvimento de um Sistema de Apoio a Decisão (SAD) para acelerar e melhorar a liberação de resultados, com economia de recursos humanos, possibilitando a diminuição de horas-extras dos funcionários da seção de Imunoquímica do Serviço de Análises Clínicas (SAC) do HC. Além disso, com a automatização dos processos, a produtividade tende a aumentar com a diminuição do tempo para a liberação dos resultados de exames, principalmente para as áreas de urgência e emergência do HC.

A pesquisa é justificada em virtude da necessidade do HC liberar o resultado de exames de pacientes com a maior rapidez possível. O presente trabalho gerará análises estatísticas possibilitando uma quantidade maior de pesquisas futuras e desenvolvimento de *softwares*.

O estudo trata ainda da aplicação de recursos tecnológicos de inteligência artificial, redes neurais artificiais, que serão aplicados referentes a exames laboratoriais clínicos para liberação de resultados baseado no conhecimento experiencial de um conjunto de dados, registrados em banco de dados, os quais colaboraram com a presente dissertação. Esta dissertação analisará padrões nos resultados de exames da população que possam vir a permitir estudos mais aprofundados e melhoria das técnicas de buscas de novas soluções usando IA.

#### 1.2.1 Metas para a Aplicação das Redes Neurais neste Projeto

O projeto possui como elemento central comparar o desempenho de redes neurais artificiais com o processo de liberação de resultados utilizando valores de corte com base em valores de referência. Embora eficaz, a liberação utilizando a valores de referência, é rígida, não flexível ao histórico do paciente ou mesmo a análise conjunta dos resultados.

As redes neuronais apresentam a flexibilidade e a possibilidade do aprendizado com os critérios vigentes e facilmente alteráveis dos profissionais, designados como customizados. Neste contexto, tem-se como hipótese central, que se as redes neuronais apresentarem bom desempenho comparado ao processo de liberação de resultados presente no laboratório, sua aceitabilidade para futuras aplicações será facilitada.

Também se tem como foco, a perspectiva de em futuro próximo, aplicar as redes neuronais a conjuntos de resultados com capacidade de predição de risco de hipertensão, síndrome metabólica e índice de controle glicêmico, entre outros.

Portanto, o estudo da aplicação das redes neuronais é justificado e sua comparação com um processo de liberação de resultados bem estabelecido (*cutoff* com valores de referência), devido ao pouco conhecimento disponível da aplicação de redes neuronais em ensaios laboratoriais. Perguntas sobre o comportamento das redes em ambiente hospitalar, com registros incompletos, diferentes tamanhos amostrais e outras variáveis necessitam de respostas somente obtidas com análises experimentais, como as propostas no presente trabalho.

## 2 OBJETIVOS

### 2.1 OBJETIVO GERAL

Estudar modelos computacionais com base em Inteligência Artificial, através do uso de redes neuronais artificiais, para classificar parâmetros bioquímicos e oferecer a possibilidade de suporte na liberação de resultados dos exames do perfil lipídico e glicemia em jejum.

### 2.2 OBJETIVOS ESPECÍFICOS

- Modelar e criar um banco de dados contendo os registros dos resultados de glicemia em jejum e perfil lipídico (colesterol total, HDL-colesterol, LDL-colesterol e triglicérides) presentes no banco de dados do HC-UFPR.
- Classificar as características extraídas dos resultados laboratoriais do perfil lipídico e glicemia em jejum com padrões baseados em valores de referência.
- Treinar e testar as redes neuronais *Multilayer Perceptron* (MLP) e *Free Associative Neurons* (FAN) e estudar o desempenho destas com o uso da classificação através de valores de referência para os exames.
- Extrair os dados divergentes da rede neuronal de melhor desempenho e submetê-los para análise de dois profissionais qualificados para identificar o motivo das divergências na classificação dos exames pela rede neuronal.
- Verificar o desempenho da rede neuronal de melhores atributos em uma base de dados considerada grande (120.000 registros com resultados de exames), mimetizando aplicação na rotina laboratorial.

### 3 REVISÃO DE LITERATURA

Com uma crescente epidemia de doenças cardiovasculares no início da década de 30, o Serviço de Saúde Pública dos Estados Unidos decidiu realizar um estudo em grande escala para investigar porque a doença cardíaca tornou-se a número um em mortes em 1940 (ROSEGREN *et al.*, 1989).

#### 3.1 O ESTUDO DE FRAMINGHAM

Um estudo iniciado em 1948 mudou a história sobre das doenças cardiovasculares. O estudo realizado em moradores na cidade Framingham, no estado de Massachusetts, EUA, tornou-se sinônimo de notáveis avanços feitos na prevenção da doença cardíaca nos Estados Unidos e em todo o mundo. Em Framingham foram selecionados pelo Serviço de Saúde Pública dos EUA 5.209 residentes saudáveis entre 30 e 60 de anos (ROSEGREN *et al.*, 1989).

Em seu primeiro ano, as responsabilidades do estudo foram assumidas pelo *National Heart Lung and Blood Institute* (NHLBI). Através de um contrato com o NHLBI, pesquisadores da Universidade de Medicina de Boston (UMB) desempenharam um papel importante no estudo do coração em habitantes de Framingham. O NHLBI e cientistas da UMB introduziram o conhecimento sobre como reduzir a morte por doença cardíaca. Essa parceria juntamente com outros colaboradores no mundo revolucionou a maneira de se ver, tratar e prevenir doenças cardiovasculares (NHLBI, 2010).

Nos cinquenta anos seguintes foram coletadas amostras de moradores de Framingham para identificar os fatores de risco associados a doenças cardíacas, derrame e outras doenças que produziram mais de 1.000 trabalhos científicos. Tal estudo fundamentou o caminho para os pesquisadores realizarem ensaios clínicos criando uma revolução na medicina preventiva, mudando a maneira como a comunidade médica e opinião pública em geral pensava, sobre a origem da doença (PHILLIPS, PINKERNELL e JING, 2004). Foi através deste estudo que se criou o Escore de Framingham que calcula o risco baixo, médio e alto do paciente desenvolver doenças cardiovasculares em 10 anos.

Antes de Framingham, os médicos acreditavam que Aterosclerose era inevitável no processo de envelhecimento e eram ensinados que a pressão arterial, que permite ao coração bombear o sangue através de artérias, deveria aumentar com a idade do paciente. Os médicos não entendiam a relação, por exemplo, entre os altos níveis de colesterol e ataques cardíacos. Não acreditavam que a modificação de determinados comportamentos poderiam evitar as causas de problemas vasculares e doenças cardíacas dos seus pacientes ou os chamados “fatores de risco” (PHILLIPS, PINKERNELL e JING, 2004).

A cada 2 ou 4 anos, participantes do estudo eram submetidos a extensos exames médicos, incluindo a história clínica, exames de sangue e outros testes mais sofisticados, como cintilografia óssea, eco cardiograma e exames oftalmológicos para avaliar vários aspectos de seu estado de saúde.

O estudo é considerado um dos mais importantes para a medicina americana. São muitas as contribuições na área de pesquisa do coração onde os pesquisadores utilizaram os dados para investigar a demência, AVC, osteoporose, artrite, diabetes, doenças oculares, câncer e os padrões genéticos de várias doenças comuns (STAMLER *et al.*, 1993).

A partir do estudo, fatores de risco para doenças cardiovasculares foram identificados. A gestão dos níveis de colesterol, pressão arterial elevada e diabetes para controlar doenças cardíacas, vasculares e derrames são fundamentais para a boa assistência médica. Antes de Framingham, o papel do colesterol sérico na evolução da doença cardiovascular não era amplamente entendido ou aceito pelos médicos como um fator de risco cardiovascular. O estudo estabeleceu uma relação entre os níveis de colesterol e risco para a doença. Além disso, o estudo estabeleceu uma forte associação positiva do colesterol da Lipoproteína de Baixa Densidade (*Low Density Lipoprotein*) com a doença arterial coronariana, bem como uma relação poderosa e inversa com efeito protetor dos níveis de colesterol da Lipoproteína de Alta Densidade (*High Density Lipoprotein*). O ganho de peso acompanhado pela falta de exercício favorece os riscos cardiovasculares, bem como hipertensão e diabetes (AUER *et al.*, 2005).

Segundo a Organização Mundial de Saúde (OMS), mais de 17,5 milhões de pessoas foram a óbito por doenças cardiovasculares e um número significativo foi a óbito por diabetes conforme Figura 1.

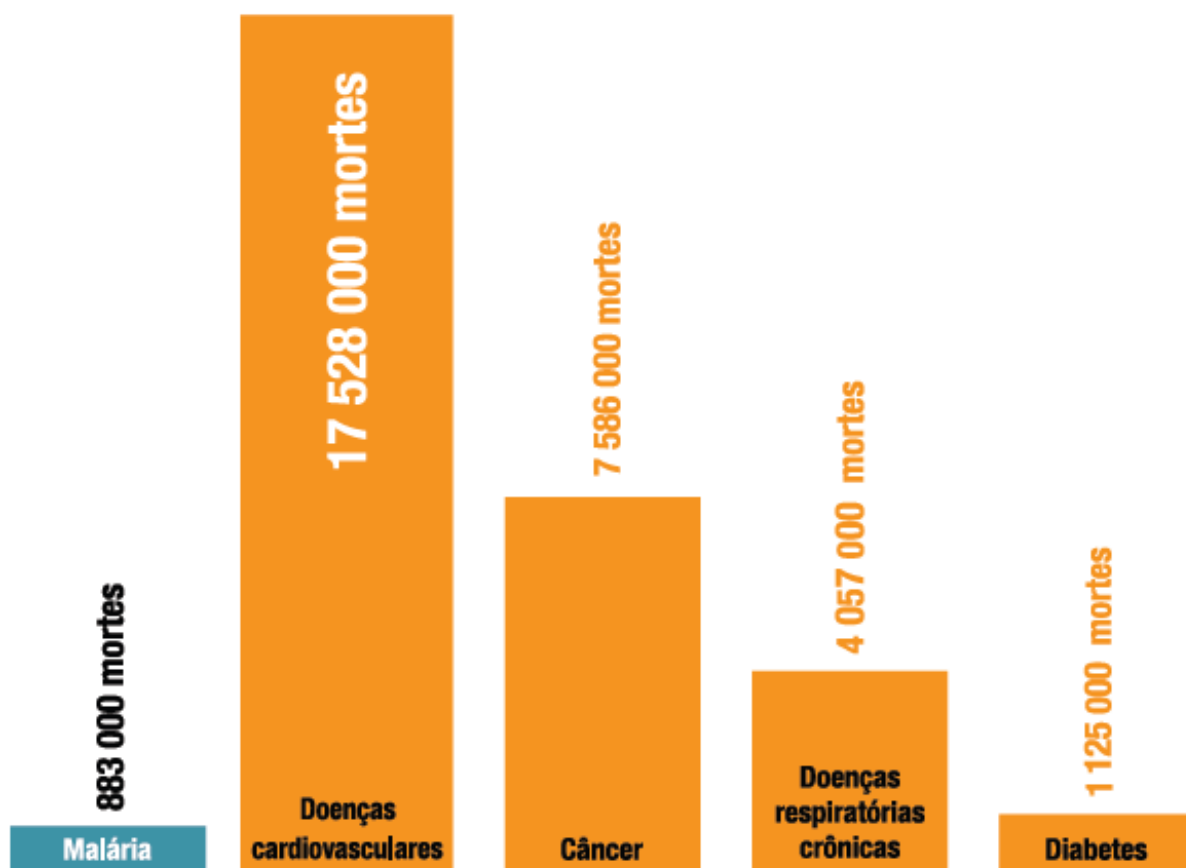


FIGURA 1 - NÚMERO DE ÓBITOS POR DOENÇA

FONTE: Organização Mundial da Saúde (OMS, 2005)

Como se pode observar, doenças cardiovasculares e diabetes levam a óbitos uma parcela considerável da população mundial. Exames de perfil lipídico que correspondem a: Colesterol Total, HDL-colesterol, LDL-colesterol, Triglicérides juntamente com Glicemia em Jejum ajudam a diagnosticar a situação da saúde da população. Esses exames servem também para prevenir a população com relação a doenças cardiovasculares e diabetes.



### 3.2 HOSPITAL DE CLÍNICAS DA UNIVERSIDADE FEDERAL DO PARANÁ

O Hospital de Clínicas (HC), um hospital público federal de ensino, é um pólo de referência no atendimento médico-hospitalar na capital do Estado do Paraná. Órgão suplementar da Universidade Federal do Paraná (UFPR), o HC tem o objetivo de servir ao ensino, a pesquisa e a extensão nas áreas das ciências da saúde e de prestar assistência à saúde da comunidade. É o maior hospital público do Paraná e o terceiro maior hospital universitário do país. O HC começou a ser construído em 1949 e foi inaugurado em 05 de agosto de 1961 pelo presidente Jânio Quadros. A Figura 2 mostra o prédio central do HC.



FIGURA 2 - HOSPITAL DE CLÍNICAS DA UFPR

FONTE: Livre adaptação de <<http://www.hc.ufpr.br>>. Acesso em 18/09/2010.

As dimensões e números do HC impressionam. Com 66.000 m<sup>2</sup> de área construída, onde circulam cerca de 10 mil pessoas diariamente, o HC possui 643 leitos, com 561 ambulatorios com 59 especialidades médicas, 3.514 funcionários, 2.276 alunos de graduação, sendo que destes, 721 são alunos de medicina, 266 professores e 261 residentes, sendo exibidos na Tabela 1 seus principais dados (HC, 2010).

TABELA 1 - ESTRUTURA GERAL DO HOSPITAL DE CLÍNICAS DA UFPR

DADOS DO HC	
Área Construída	66.000 m²
Total de Funcionários	3.514
Ambulatórios	561
Leitos	643
Cirurgias (mensal)	716
Especialidades Médicas	59

FONTE: Hospital de Clínicas da UFPR (HC, 2010)

O ensino de graduação é acompanhado pelos Departamentos de Clínica Médica, Cirurgia, Otorrinolaringologia e Oftalmologia, Saúde Comunitária, Tocoginecologia, Enfermagem, Terapia Ocupacional e Nutrição. Sendo um dos maiores hospitais universitários do país com vários serviços de excelência e títulos que conquistou ao longo destes anos. Entre eles estão:

- Banco de Leite Humano do HC criado em 1978 e referência no Paraná, segundo a Secretaria Estadual da Saúde;
- Hospital de Excelência como o Serviço de Transplante de Medula Óssea, criado em 1979, referência internacional e pioneira na América Latina nos transplantes entre não aparentados e de células de sangue de cordão umbilical;
- HC recebeu do Fundo das Nações Unidas para a Infância (UNICEF) em dezembro de 1995, o título "Hospital Amigo da Criança";
- No ano de 2009 foi referenciado para atender, prioritariamente, os casos de gripe influenza A H1N1;
- Outros serviços que são destaque: Oftalmologia - Banco de Olhos, Transplantes Hepático, Cardíaco e Renal, Neurologia e o Diagnóstico de Doenças Neuromusculares, Cirurgia de Epilepsia, Terapia Intensiva Neonatal, Farmácia Hospitalar, Ortopedia e Traumatologia.

### 3.2.1 Unidade de Apoio Diagnóstico

A Unidade de Apoio Diagnóstico (UAD) do HC-UFPR é responsável em auxiliar no diagnóstico de doenças e promover as terapias equivalentes em razão da grande necessidade de automatização de processos nas diversas seções dos laboratórios do Serviço de Análises Clínicas (SAC).

A UAD possui várias seções executando em média 105 mil exames por mês e atendendo cerca de 20 mil pacientes por mês. Agrega dois serviços, o de Análises Clínicas, que atua nas áreas de Biologia Molecular, Bioquímica, Imunologia, Hormônios, Imunogenética, Bacteriologia, Micologia, Parasitologia, Hematologia e Urinálise; e o de Anatomia Patológica, que realiza diagnósticos de alto padrão a partir das biópsias, que resultam no diagnóstico definitivo para a grande maioria das neoplasias e outros grupos de doenças. Possui uma equipe com cerca de 270 profissionais, entre farmacêuticos, biólogos, médicos, enfermeiros, técnicos e auxiliares de laboratório e funcionários administrativos (HC, 2010).

No ano de 2009 a Unidade teve sua atenção voltada à infraestrutura e condições de trabalho dos profissionais, visando à melhoria da qualidade dos serviços prestados. Nesse sentido, realizou uma ampla reforma no segundo andar do laboratório, abrangendo todas as seções, unificando áreas como a Bioquímica, Hormônios e Sorologia, o que resultou na eliminação de tarefas repetidas, diminuição do tempo de liberação dos resultados de exames e redução de custos.

A Seção de Imunoquímica possui 32 funcionários trabalhando em regime de funcionamento contínuo, realiza em média 90 mil exames por mês, distribuídos em 86 testes analíticos diversos de alta e média complexidade. Com relação aos exames específicos do perfil lipídico e glicemia em jejum, objeto da dissertação, a seção realiza em média 15 mil exames por mês.

Para realizar a grande quantidade de exames, a seção conta com vários equipamentos laboratoriais, entre eles, existem dois equipamentos da Abbott® chamado Architect ci8200 conforme Figura 3 e que possui autonomia de produzir até 1,2 mil exames por rodada (ABBOTT, 2010).



FIGURA 3 - ANALISADOR AUTOMÁTICO ARCHITECT ci8200

FONTE: Livre adaptação de <<http://www.mustf-miar.org.mo>>. Acesso em 11/11/2010

A UAD participa do Programa Nacional de Indicadores Laboratoriais (ControlLab) da Sociedade Brasileira de Patologia Clínica, que compara diversas atividades entre os laboratórios e envia, trimestralmente, relatórios com o posicionamento de cada estabelecimento, o que possibilita, a partir da análise dos indicadores, desenvolver estratégias para ações de melhoria.

### 3.2.2 Unidade de Informação

A unidade de informação foi criada no início de 1990 com o objetivo de coordenar o desenvolvimento de sistemas que permitissem integrar todos os setores envolvidos na produção, armazenamento e utilização de informações no HC. Dando prioridade à melhoria da qualidade do atendimento e ao funcionamento do hospital, sua primeira tarefa foi o desenvolvimento do Sistema de Informações Hospitalares (SIH).

Devido às necessidades de melhoria das atividades de ensino e a consolidação das atividades de pesquisa desenvolvidas no HC, tarefas adicionais foram também contempladas, incluindo a informatização dos departamentos de ensino, o apoio ao preparo e apresentação de trabalhos científicos, o treinamento de usuários em informática e a integração à rede de computadores da UFPR.

Desde o início da década de 90 a equipe de informática trabalha na informatização do HC, desenvolvendo o SIH que é um sistema com arquitetura centralizada. Vários módulos (subsistemas) integram o SIH, todos desenvolvidos especificamente para a área médico-hospitalar.

O SIH possui 18 módulos implantados, com cerca de 700 tabelas no banco de dados DB2® e 3 mil aplicações, para dar suporte aos serviços de atendimento a pacientes e serviços administrativos. Foi originalmente desenvolvido utilizando-se a linguagem *Cross Structured Program* (CSP) e posteriormente substituído pelo *Visual Age Generator*, ambos da IBM®. O SIH era executado num computador *mainframe* IBM S/390, com banco de dados DB2, funcionava em terminais IBM 3270 e microcomputadores com emulação de terminal.

Os dados dos pacientes do HC foram disponíveis para administradores, médicos e pesquisadores e todos os módulos possuem relatórios com informações gerenciais e de controle. Porém algumas vezes são necessárias informações para as quais não foram previstas funções para recuperá-las no sistema. Como não existe modo de prever todos os dados que os usuários necessitam a cada momento para gestão ou para pesquisa, utilizava-se o *Query Management Facility* (QMF), *software* da IBM que permitia a construção de consultas para recuperar informações gerenciais e estatísticas. No momento utiliza-se o editor de comandos do DB2.

Seja para fazer manutenções ou para obter informações é necessário conhecimento do modelo de dados, suas relações e condições de integridade. Isto não é uma tarefa fácil e rápida para quem não conhece o sistema. Sem ajuda de outras pessoas, sem documentação, sem uma metodologia e sem uma ferramenta que ajude a obter informações do código isso se transforma numa tarefa complexa e demorada, levando um tempo que normalmente os usuários não possuem.

Não se pode ficar alheio ao real benefício que a informática trouxe ao campo da medicina, pois a quantidade de informações que podem ser arquivadas e correlacionadas é útil nos estudos estatísticos e epidemiológicos bem como para o gerenciamento das unidades funcionais de um hospital.

A criação do SIH visa conciliar atendimento e gerenciamento, tornando mais dinâmico o acesso às informações geradas no ambiente médico-hospitalar.

O HC possui grande experiência em informática hospitalar e foi um dos pioneiros em informatização de hospitais de grande porte no país, tendo desenvolvido o sistema transacional de atendimento que está em uso na instituição.

O SIH foi processado em computador IBM de grande porte com banco de dados DB2 até o ano de 2006 quando ocorreu uma migração dos sistemas e a instituição substituiu o *mainframe* S/390 por servidor *Advanced Interactive eXecutive* (AIX). O sistema conta com diversos módulos para controle e gerenciamento das funções de atendimento do hospital abrangendo serviços de atendimento ambulatorial, internação, exames laboratoriais e complementares, prescrição médica, enfermagem, banco de sangue, além de módulos administrativos e de apoio, tais como faturamento, controle de medicamentos e material hospitalar. O sistema opera 24 horas por dia e 365 dias por ano.

No entanto, uma de suas limitações refere-se à dificuldade e, eventualmente, a impossibilidade de cruzamento de dados pertencentes a módulos diferentes. Esta limitação impede a realização de comparações ou mesmo a obtenção de importantes informações de ordem processual anteriormente não previstas nos relatórios fornecidos pelo sistema. Essa insuficiência de agilidade na evolução do sistema deve-se normalmente à pressão sobre custo, à rotatividade de pessoal, ao tempo necessário para os novos desenvolvimentos e à falta de documentação própria do sistema. Ao longo dos anos, todos os módulos sofreram modificações para acrescentar e modificar funcionalidades que não foi documentada. Os usuários e projetistas que programaram as modificações já não são mais os mesmos. Alguns módulos já sofreram tantas alterações que cada vez fica mais difícil fazer novas manutenções. Além disso, o número de usuários aumentou na medida em que os novos módulos foram implantados, o número de analistas e programadores diminuiu e a demanda por novos módulos e funcionalidades aumentou.

Com a intenção de acompanhar a evolução dos ambientes computacionais e permitir a absorção de novas tecnologias, o HC está se atualizando tecnologicamente. A base de dados do SIH será incorporada por um sistema em ambiente *web*. A proposta de atualização tecnológica inclui também a adoção de padrões internacionais de interfaceamento de componentes de *software*.

Um módulo do SIH importante é o de Laboratório, pois cerca de 40% do processamento do servidor é ocupado por este módulo. O médico ou profissional que necessita requisitar exames o pode fazer em qualquer lugar do HC ou até remotamente o faz, desde que tenha acesso ao mesmo (FIGURA 4).

```

HC-UFPR                                REQUISICAO DE EXAMES                                LB75M01
-----
  Cliente : 13864985 - ADEMIR LUIZ DO PRADO
  Requisicao :                               Laboratorio : 1 - CENTRAL
  Solicitante : ADMAR - ADMAR MORAES DE SOUZA
  Ind.Clinica : COLESTEROL                      Agenda coleta : N

  St      Exame
  (C)     Codigo  Descricao
  --      -
      285 - COLESTEROL TOTAL
      270 - COLESTEROL HDL
      271 - COLESTEROL LDL
      280 - TRIGLICERIDEOS
      [ ] -
      -
      -

  Amostra
  Codigo  Descricao
  --      -
      2 - SANGUE
      2 - SANGUE
      2 - SANGUE
      2 - SANGUE
      -
      -

  Urg
  --
      S
      S
      S
      S
      S
      S

PF:1-Ajuda    2-Busca    4-Altera    5-Exclui    7-Volta Pg  8-Avanca Pg
10-Registra
LBMS0199I    AO FINAL TECLE PF10 PARA REGISTRAR
  
```

FIGURA 4 - EXEMPLO DE REQUISIÇÃO INFORMATIZADA DE EXAMES  
LABORATORIAIS NO SIH

FONTE: Hospital de Clínicas da UFPR

Uma vez requisitado exames, se o paciente estiver internado no HC, um técnico de laboratório irá até o leito aonde se encontra o paciente e coletará amostra. No caso de paciente ser atendido em ambulatório, o mesmo irá ao setor de coleta de exames no dia e hora marcados.

Após a coleta de amostra, os exames são encaminhados para a seção responsável pela realização dos resultados. Neste setor os exames são submetidos ao equipamento chamado Architect ci8200 e após análise é emitido resultado através de um relatório que é digitado no SIH (FIGURA 5).

HC-UFPR	REGISTRO DE RESULTADO DE DETERMINACOES		LBD6MO1
Lab.: 1 - CENTRAL	Secao: 35 - IMUNOQU.1	Lista: 8376	Data: 22 / 09 / 2010
Nro.: 6	Cliente : 13864985 - ADEMIR LUIZ DO PRADO	Sx : M	
	Idade : 40a00m	Ind.Clin.:	
	Exame : 285 - COLEST	Amostra : 2	- SANGUE
	Metodo : 134 - ENZIMATICO/COLORIMETRICO		
COLEST : 200			
PF:1-Ajuda 2-Busca 5-Padiao 7-Volta Pg 8-Avanca Pg 9-Libera 10-Registra 12-Retorna LBMS0210I EXISTEM MAIS DETERMINACOES - TECLE PF8			

FIGURA 5 - EXEMPLO DE REGISTRO DE RESULTADO INFORMATIZADO DE EXAMES LABORATORIAIS NO SIH

FONTE: Hospital de Clínicas da UFPR

No HC-UFPR estão informatizados e são registrados 423 tipos de exames laboratoriais clínicos. No ano de 2010 foram realizados 1.462.287 exames que foram registrados no sistema de informática do HC. Deste número, 168.643 correspondem a apenas 5 exames: Colesterol Total, HDL-colesterol, LDL-colesterol, Triglicérides e Glicemia em Jejum. Apenas estes 5 exames correspondem a 11,5% do total geral.

Sistemas de Informações Hospitalares são utilizados usualmente para gerenciar hospitais. Além de sistemas informatizados convencionais, existe ainda o diagnóstico com auxílio de computador que pode ser definido como um diagnóstico realizado por um profissional, por exemplo, um radiologista, que utiliza o resultado de análises quantitativas automatizadas de imagens radiográficas processadas por programas computacionais como uma segunda opinião para a tomada de decisões diagnósticas (HAWKER, 2007).

O objetivo do *Computer-Aided Diagnosis* (CAD) ou Diagnóstico Auxiliado por Computador é acertar o diagnóstico, como a interpretação da imagem radiológica, mediante o uso da resposta do computador como referência (QUINLAN, 1986). O CAD não substitui o profissional e sim serve como apoio para uma segunda opinião.



### 3.3 PERFIL LIPÍDICO E GLICEMIA EM JEJUM

#### 3.3.1 Glicemia em Jejum

A glicemia é a concentração de glicose no sangue, mais precisamente no plasma. O corpo transforma alguns dos carboidratos que comemos em glicose e a glicemia é o nível de glicose presente no sangue. Quando uma pessoa se alimenta muito, a glicemia aumenta, ao passo que quando come pouco, ela se mantém baixa. Mede-se a glicemia através da confirmação dos sinais e sintomas clássicos da glicemia em jejum (exame de sangue onde são verificadas as taxas de glicose no sangue) e do teste padronizado de Teste Oral de Tolerância à Glicose (TOTG) (TAKAISHI *et al.*, 2004).

Para tolerância à glicose em jejum de 111 a 125 mg/dL e 2 horas após 75 g de glicose os valores são de 141 a 199 mg/dL. Para o Diabetes *mellitus* em jejum o valor maior que 126 mg/dL e 2 horas após 75 g de glicose os valores são maior que 200 mg/dL. Os critérios de diagnóstico estão baseados nas recomendações da comunidade científica (ADA, 2010).

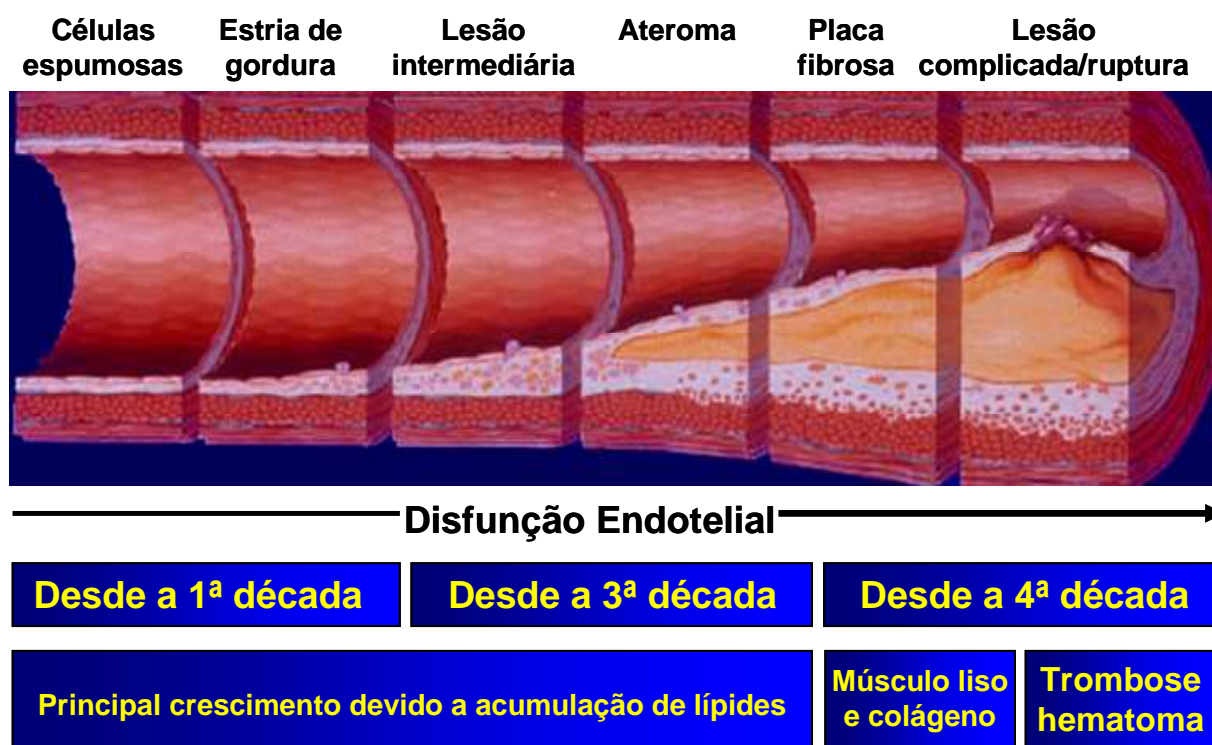
Além do uso da insulina, diabéticos podem controlar a glicemia através de dietas específicas e prática de exercícios físicos. A prática regular de exercício físico aumenta a ação da insulina, fazendo com que a glicose saia da corrente sanguínea, diminuindo a glicemia. Uma boa taxa de glicemia seria entre 70 e 100 mg/dL (SBC, 2007).

#### 3.3.2 Perfil Lipídico

O Perfil Lipídico é composto pelas determinações do Colesterol Total (CT), colesterol ligado à HDL-c, colesterol ligado à LDL-c e Triglicérides (TG). Lipídeos são substâncias de origem orgânica, caracterizadas pela insolubilidade em água e solubilidade em benzeno, éter e clorofórmio.

O colesterol é o esteroide mais abundante nos tecidos humanos, derivado do ciclopentanoperidrofenantreno com ligações duplas entre C-5 e C-6, hidroxila no C3 (colesterol livre) ou ácido graxo no C3 (colesterol esterificado) e cadeia alifática de 8 carbonos no C17. As funções do colesterol no organismo humano são: precursor dos hormônios esteróides, dos ácidos biliares e da vitamina D, como constituinte das membranas celulares que ajuda na fluidez e ativação das enzimas (MOTTA, 2009).

A Figura 6 mostra o processo de desenvolvimento da placa aterosclerótica que se desenvolve a partir da deposição do colesterol na íntima das artérias.



**FIGURA 6 - EVOLUÇÃO DO PROCESSO ATEROSCLERÓTICO**

A deposição de colesterol na íntima das artérias leva ao processo de aterosclerose que se desenvolve de forma lenta e progressiva.

FONTE: Adaptado de Stary HC *et al. Circulation.* 1995,;1355-1374

Os valores da Tabela 2 foram utilizados como referência para os valores séricos dos componentes do perfil lipídico.

**TABELA 2 - VALORES DE REFERÊNCIA PARA O PERFIL LIPÍDICO**

VALORES DE REFERÊNCIA PARA ADULTOS (> 19 anos) em mg/dL			
EXAMES	DESEJÁVEIS	LIMÍTROFES	ALTERADO*
Colesterol Total	Abaixo de 200	200-239	Acima de 240
LDL colesterol	100-129	130-159	Acima de 160
HDL colesterol	Acima de 40 (Homens) Acima de 50 (Mulheres)		Abaixo de 40 (Homens) Abaixo de 50 (Mulheres)
Triglicérides	Abaixo de 150		Acima de 150

\* Riscos de problemas cardiovasculares

FONTE: III Diretrizes Brasileiras em Dislipidemia (SBC, 2001)

O colesterol é sintetizado pelos tecidos hepático e extra-hepático em um processo regulado por um sistema compensatório: quanto maior for a ingestão de colesterol vindo dos alimentos, menor é a quantidade sintetizada pelo tecido.

O colesterol é um lipídeo e por isso é insolúvel em água e, conseqüentemente, insolúvel no sangue. Para ser transportado na corrente sanguínea, o colesterol liga-se com proteínas, chamadas de apolipoproteínas, e outros lipídeos como Triglicérides e Fosfolipídeos em um complexo chamado de Lipoproteína (MOTTA, 2009).

Existem quatro classes de lipoproteínas separadas em dois grupos: primeiro, as ricas em triglicérides, maiores e menos densas, representadas pelos Quilomicrons de origem intestinal, e pelas lipoproteínas de densidade muito baixa ou *Very Low Density Lipoprotein* (VLDL), de origem hepática; e segundo, as ricas em colesterol de densidade baixa ou *Low Density Lipoprotein* (LDL) e de densidade alta ou *High Density Lipoprotein* (SCARTEZINI *et al.*, 2003).

Entre as lipoproteínas de interesse na clínica estão as LDL e HDL. A lipoproteína LDL representa a maior fração das lipoproteínas aterogênicas e, portanto, seus níveis elevados no sangue estão associados com a Doença Arterial Coronariana (DAC). A lipoproteína HDL tem função protetora que é a de transportar o excesso de colesterol dos tecidos para o fígado para a síntese de ácidos biliares ou ser eliminado na secreção biliar. Esse processo é chamado de “transporte reverso do colesterol” e tem importância por evitar a formação de aterosclerose que é um fator determinante para a DAC (PHILLIPS, PINKERNELL e JING, 2004).

As concentrações dessas lipoproteínas, LDL e HDL, são expressas em laboratório clínico pela quantificação do colesterol da lipoproteína como LDL-colesterol e HDL-colesterol. As concentrações de colesterol-lipoproteína foram determinadas na maior parte dos estudos populacionais relacionando a concentração das lipoproteínas com o risco cardiovascular (ROSEGREN *et al.*, 1989).

A dosagem do colesterol total, que foi absoluta até a década de 70, representa a concentração de colesterol na corrente circulatória distribuída nas principais lipoproteínas circulantes: VLDL, LDL e HDL. O fracionamento do colesterol total é traduzido como a somatória das determinações laboratoriais VLDL-C + LDL-C + HDL-C (SCARTEZINI *et al.*, 2003).

Os triglicérides são os principais lipídeos originários da alimentação, transportados pelos Quilomicrons do intestino ao fígado, mas são sintetizados pelo fígado e outros tecidos no organismo humano. A lipoproteína VLDL produzida pelo fígado é a responsável pelo transporte dos triglicérides no sangue. Níveis elevados de triglicérides (acima de 200 mg/dL) acompanham a elevação da lipoproteína VLDL no sangue e estão associados à maior ocorrência de DAC. Explica-se, em parte, pelos baixos níveis de lipoproteína HDL quando os níveis de Triglicérides estão altos no sangue (PICHETH *et al.*, 2001).

A ingestão de excesso de gordura e bebida alcoólica pode elevar a concentração sérica dos triglicérides, fator predisponente para pancreatite (inflamação do pâncreas) e fígado gorduroso (esteatose hepática) (PICHETH *et al.*, 2001).

### 3.4 INFORMÁTICA E O LABORATÓRIO DE ANÁLISES CLÍNICAS

As tecnologias da informação e comunicação, aliadas à ciência médica, têm contribuído para a melhoria dos sistemas de saúde. É impraticável gerenciar um laboratório de análises clínicas sem o uso da informática.

Para facilitar a agilidade da informação são necessários que diferentes sistemas informatizados possam se comunicar entre si. Para isso surge a necessidade de interfaceá-los. No HC foi desenvolvido um sistema que integra o SIH com os *softwares* que se comunicam com os equipamentos laboratoriais. Para interfacear o sistema de laboratório do HC foi necessário o desenvolvimento e implantação do recurso do código de barras nos tubos dos exames (FIGURA 7).



**FIGURA 7 - EXEMPLO DE TUBO DE COLETA DE SANGUE COM ETIQUETA DE CÓDIGO DE BARRAS PARA INTERFACEAMENTO**

O tubo para coleta de sangue mostra a estrutura do código de barras utilizado no Hospital de Clínicas da UFPR. A etiqueta possui informações do paciente como número do registro e nome, informações dos exames tais como descrição da amostra, descrição dos exames solicitados bem como número e data/hora da coleta da amostra.  
 FONTE: O autor (2011)

Foi um pré-requisito para desenvolver o sistema de interfaceamento, haja vista ser necessário à comunicação entre o equipamento que libera os resultados de exames e a persistência (gravação) final dos resultados dos exames no SIH. A etiqueta foi desenvolvida pelo pesquisador no sistema SIH para viabilizar o interfaceamento no HC.

Para melhor visualizar, veja abaixo na Tabela 3, qual a diferença entre um laboratório que possui interfaceamento e um laboratório que não possui.

**TABELA 3 - EFEITO DO INTERFACEAMENTO EM LABORATÓRIOS DE ANÁLISES CLÍNICAS**

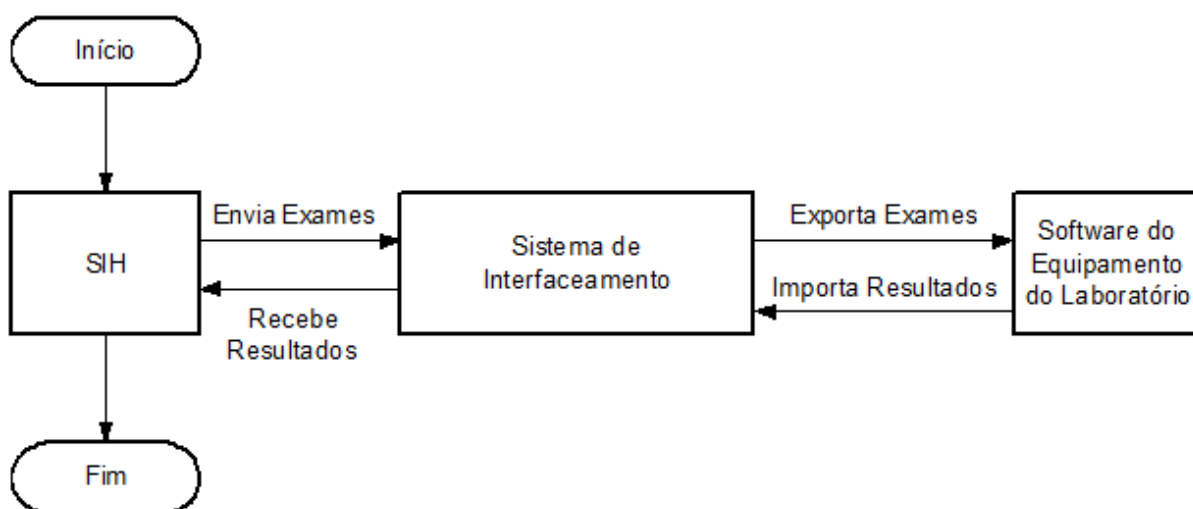
<b>CARACTERÍSTICAS DOS LABORATÓRIOS CLÍNICOS</b>		
<b>Processos</b>	<b>Sem interfaceamento</b>	<b>Interfaceado</b>
Cadastro de exames nos analisadores automáticos	Falta cadastrar um exame no equipamento.  Cadastra o exame errado.	O computador não “esquece” de cadastrar nenhum exame no equipamento.  Cadastro corretos de todos os exames.
Introdução das amostras no equipamento	Troca os tubos das amostras dos pacientes.	Impossível ocorrer troca de tubos de amostras, pois todos os tubos são marcados por código de barras.
Transcrição da cópia do resultado do equipamento para a ficha do paciente	Cópia do resultado de outro paciente.	Com o interfaceamento esta etapa é eliminada.
Digitação do resultado do exame	Possibilidade de troca de resultados entre pacientes.  Erro de digitação.	O resultado é eletronicamente associado ao paciente.  Não há erros de digitação.

FONTE: Estudo próprio no Hospital de Clínicas da UFPR (HC, 2010)

Após constatação acima das vantagens do interfaceamento, foi desenvolvido um *software* para atender a necessidade de comunicação entre o sistema que libera resultados dos equipamentos do laboratório e o SIH. O profissional contratado pelo HC em 2010 em parceria com o pesquisador atendeu a uma necessidade importante do laboratório de análises clínicas do HC. A expectativa dos profissionais de laboratório do HC pela implantação do interfaceamento ultrapassava a uma década.

#### 3.4.1 Interfaceamento no Hospital de Clínicas

O *software* desenvolvido permitiu a integração entre sistemas de informática. Um sistema responsável por gerar os resultados e outro responsável por receber os resultados dos exames do paciente. A Figura 8 apresenta um fluxograma de como funciona o Sistema de Interfaceamento e sua integração com o módulo já existente do SIH descrito no item 3.2.2 e o *software* do equipamento laboratorial.



**FIGURA 8 - FLUXOGRAMA DAS ETAPAS DO INTERFACEAMENTO**

O médico solicita exames laboratoriais através do SIH. O sistema de interfaceamento envia os exames da requisição para o *software* do equipamento. O *software* do equipamento envia para o sistema de interfaceamento. O sistema de interfaceamento grava os resultados dos exames no SIH. O médico consulta/imprime os resultados.

FONTE: O autor (2011)

O Sistema de Interfaceamento possibilita a integração do SIH, onde o médico consulta e imprime resultados de exames, com o *software* dos equipamentos laboratoriais que emite os resultados de exames.

O sistema é executado dentro da intranet do HC e o usuário necessita *logar* no sistema (FIGURA 9) para monitorar a exportação dos exames laboratoriais e a importação de resultados de exames.



FIGURA 9 - EXEMPLO DE INTERFACE COM LOGIN DE USUÁRIO NO SISTEMA DE INTERFACEAMENTO

FONTE: Hospital de Clínicas da UFPR

A importação e exportação dos exames do SIH para o equipamento que realizará os exames são gerenciadas pelo usuário *logado*. O usuário poderá parar ou iniciar a importação / exportação dos dados (FIGURA 10), “limpar” possíveis erros identificados no processo que possam ocorrer e que são gravados em arquivos de *log* gerados da movimentação dos exames no sistema.

Um arquivo de *log* tem a função de auditoria do sistema. Toda a movimentação de exames ocorrida entre o SIH e o *software* conectado ao equipamento é monitorada nos dois sentidos: no envio dos exames e no retorno dos resultados dos exames enviados pelo *software* do equipamento.

O sistema possibilita gerenciar o tempo que se deseja enviar exames e receber os resultados dos exames. Esse tempo varia, geralmente, a cada 2 minutos para exportar e importar. O usuário *logado* deverá observar o campo “Erros” no monitoramento para tomar atitude de verificar se há conexão com o banco de dados do SIH, por exemplo.

**HOSPITAL DE CLÍNICAS**  
UNIVERSIDADE FEDERAL DO PARANÁ

Opções Sistema Sair

Status de Inicialização do Sistema

**Status de Inicialização do Sistema**

**Status de Exportação de Requisições**

Exportação	INICIADA
No Momento	AGUARDANDO
Erros	NÃO HÁ ERROS
Comunicação	BANCO ONLINE.

Exportando... Parar Exportação Limpar Erros

**Status de Importação de Requisições**

Importação	INICIADA
No Momento	EM EXECUÇÃO
Erros	NÃO HÁ ERROS
Comunicação	BANCO ONLINE.

Importando... Parar Importação Limpar Erros

HC707 Serviço de Informática do HC-UFPR - Ramal 6224

**FIGURA 10 - EXEMPLO DE INTERFACE DO PROCESSO DE IMPORTAÇÃO E EXPORTAÇÃO DE EXAMES DO SISTEMA DE INTERFACEAMENTO**  
FONTE: Hospital de Clínicas da UFPR

O sistema recebe os resultados enviados pelo *software* do equipamento. Por segurança há sempre um profissional de laboratório para liberar os exames analisando-os e tomando a decisão de liberar o resultado ou ainda submeter à amostra novamente ao equipamento para realizar nova análise da amostra.

Para desenvolvimento deste sistema o HC utilizou *softwares* livres como *Java* para definir as regras do negócio, *Hibernate* para persistir (gravar) os dados, *Ajax* para tratamento na interface através de mensagens informativas, *JBoss* como servidor *web*, *Eclipse* como o ambiente de desenvolvimento do sistema e o *Framework Click* para desenhar a interface para o usuário, entre outras tarefas.



### 3.5 INTELIGÊNCIA ARTIFICIAL

A inteligência artificial (IA) é um ramo da ciência da computação que se propõe a elaborar dispositivos que simulem a capacidade humana de raciocinar, perceber, tomar decisões e resolver problemas (RUSSELL e NORVIG, 2009).

Existente há décadas, esta área da ciência é grandemente impulsionada com o rápido desenvolvimento da informática e da computação, permitindo que novos elementos sejam rapidamente agregados a IA. Iniciada na década de 40, a pesquisa em torno desta ciência era desenvolvida apenas para procurar novas funcionalidades para o computador. Com o advento da Segunda Guerra Mundial, surgiu também a necessidade de desenvolver a tecnologia para impulsionar a indústria bélica.

Com o passar do tempo, surgem várias linhas de estudo da IA, uma delas é a biológica, que estuda o desenvolvimento de conceitos que pretendiam imitar as redes neuronais humanas. Na verdade, é nos anos 60 em que esta ciência recebe definição de inteligência artificial e os pesquisadores da linha biológica acreditavam ser possíveis máquinas realizarem tarefas humanas complexas, como raciocinar (RUSSELL e NORVIG, 2009).

#### 3.5.1 Reconhecimento de Padrões

A inteligência artificial utiliza-se do Reconhecimento de Padrões (RP) para analisar determinado conjunto de dados ou conjunto de treinamento e organizá-los de acordo com padrões. O RP visa classificar dados baseados em conhecimento preliminar ou dedutivo ou ainda informações estatísticas extraídas de padrões.

Segundo Haykin (2001), o reconhecimento de padrões é formalmente definido como o processo pelo qual um padrão ou sinal recebido é atribuído a uma classe dentre um número predeterminado de classes. É um processo de identificação de estruturas nos dados por comparações com estruturas conhecidas. Seu objetivo principal é a classificação. A abordagem comum em RP é dividir o problema em duas partes: extração das características e classificação (KUSTER *et al.*, 2006).

Estes padrões a serem classificados, normalmente, são grupos de medidas ou observações que definem pontos em um espaço multidimensional apropriado. Antes de partir para a análise efetiva, uma etapa de treinamento é realizada: nela o algoritmo de reconhecimento é testado para que seja possível saber se ele encontra os resultados esperados. Conforme RAITTZ (2002), na prática computacional, o RP pode-se enumerar algumas áreas em que vem sendo usado:

- Processamento de imagens;
- Classificação/análise dos sinais de radar;
- Reconhecimento de fisionomia;
- Reconhecimento de voz;
- Tratamento de impressões digitais;
- Reconhecimento de caracteres;
- Análise de manuscritos;
- Análise de eletrocardiogramas e eletro encefalogramas;
- Diagnósticos médicos.

Existem dois tipos de reconhecimento: o supervisionado e o não supervisionado, porém ambos entram em ação após a identificação do padrão. O reconhecimento supervisionado utiliza o conjunto de treinamento para classificar os dados obtidos de acordo com as categorias já existentes e nelas organizá-los. Por sua vez, o reconhecimento não supervisionado utiliza o conjunto de treinamento para criar novas categorias, ao invés de simplesmente separar os dados de acordo com as categorias já existentes (BISHOP, 1995).

### 3.5.2 Redes Neurais Artificiais

É do conhecimento da comunidade científica que o cérebro é composto de bilhões de neurônios. Um neurônio é uma célula formada por três seções com funções específicas e complementares: corpo, dendritos e axônio. Os dendritos captam os estímulos recebidos em um determinado período de tempo e os transmitem ao corpo do neurônio, onde são processados. Quando tais estímulos atingirem determinado limite, o corpo da célula envia novo impulso que se propaga pelo axônio e é transmitido às células vizinhas por meio de sinapses.

Este processo pode se repetir em várias camadas de neurônios. Como resultado, a informação de entrada é processada, podendo levar o cérebro a comandar reações físicas (FERNEDA, 2006). A Figura 11 ilustra de forma simplificada as partes de um neurônio biológico.

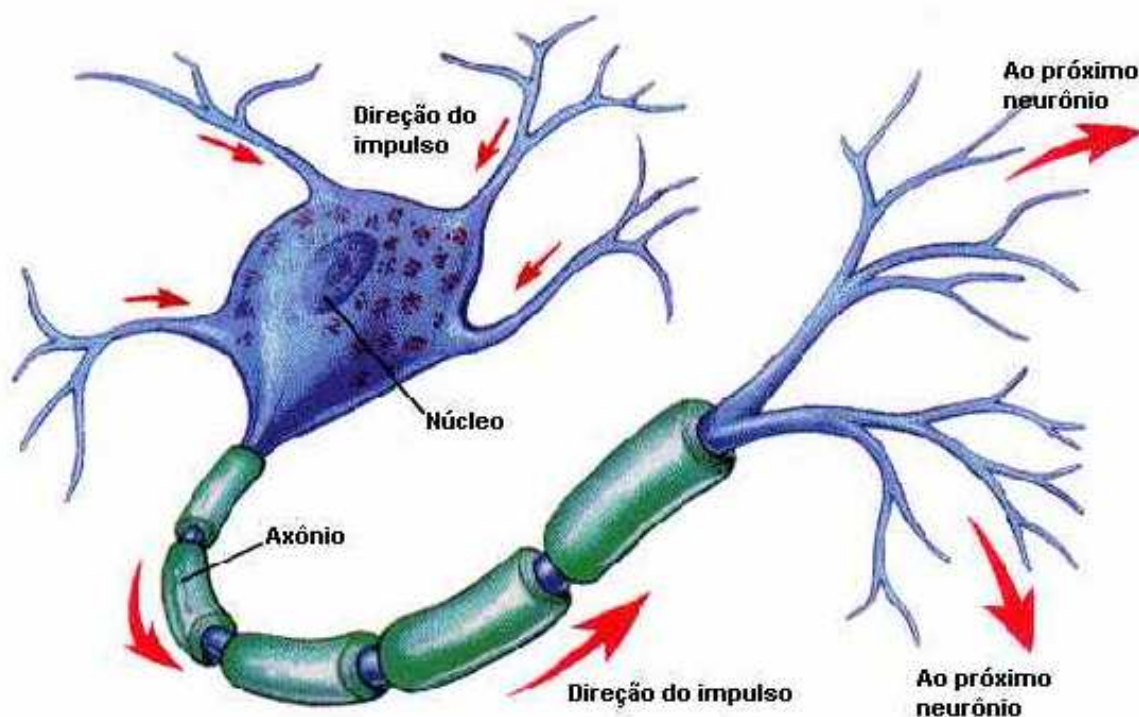


FIGURA 11 - REPRESENTAÇÃO SIMPLIFICADA DE UM NEURÔNIO BIOLÓGICO

FONTE: Livre adaptação <<http://www.lookfordiagnosis.com>>. Acesso em 18/01/2011.

A habilidade de um ser humano em realizar funções complexas e principalmente a sua capacidade de aprender advém do processamento paralelo e distribuído da rede de neurônios do cérebro. Os neurônios do córtex, a camada externa do cérebro, são responsáveis pelo processamento cognitivo. Um novo conhecimento ou uma nova experiência pode levar a alterações estruturais no cérebro. Tais alterações são efetivadas por meio de um rearranjo das redes de neurônios, reforçando ou inibindo algumas sinapses (HAYKIN, 2001). De modo semelhante trabalham as Redes Neurais Artificiais, ou seja, através de um processo de aprendizagem (BRAGA, CARVALHO e LUDEMIR, 2000).

A busca por um modelo computacional que simule o funcionamento das células do cérebro data dos anos 40, com o trabalho de McCulloch e Pitts (1943). O entusiasmo pela pesquisa neste campo cresceu durante os anos 50 e 60. Nesse período, Rosenblatt (1958) propôs um método inovador de aprendizagem para as RNAs denominado *Perceptron*.

Até 1969, muitos trabalhos foram realizados utilizando o *Perceptron* como modelo. No final dos anos 60, Minsky e Pappert (1969) publicaram um livro no qual apresentavam importantes limitações do *Perceptron*. As dificuldades metodológicas e tecnológicas, juntamente com os ataques extremamente pessimistas de Minsky e Pappert, fizeram com que as pesquisas arrefecessem nos anos seguintes. Durante os anos 70, a pesquisa contava apenas com um número ínfimo de cientistas. Porém, durante os anos 80, o entusiasmo ressurgiu graças a avanços metodológicos importantes e ao aumento dos recursos computacionais disponíveis. O modelo de neurônio artificial da Figura 12 é uma simplificação do modelo apresentado por Haykin (2001).

#### Entradas e Pesos

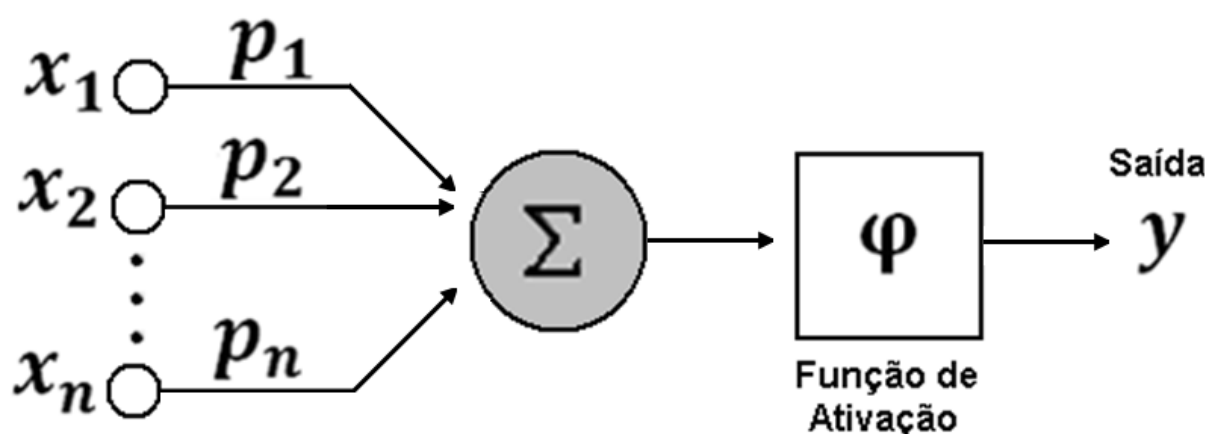


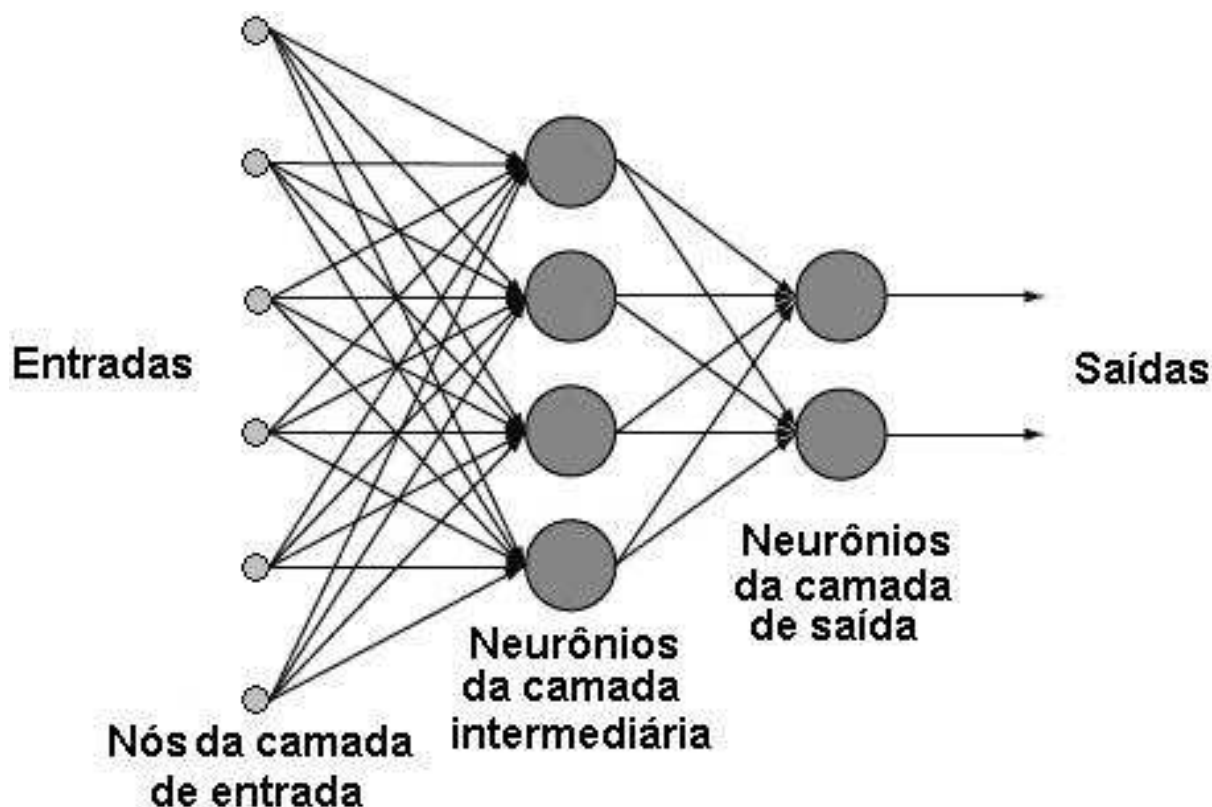
FIGURA 12 - ESTRUTURA ESQUEMÁTICA DO MODELO MATEMÁTICO DE UM NEURÔNIO

Este modelo é composto por três elementos básicos: um conjunto de  $n$  conexões de entrada ( $x_1, x_2, \dots, x_n$ ), caracterizadas por pesos ( $p_1, p_2, \dots, p_n$ ); um somador ( $\Sigma$ ) para acumular os sinais de entrada; uma função de ativação ( $\phi$ ) que limita o intervalo permissível de amplitude do sinal de saída ( $y$ ) a um valor fixo.

FONTE: Adaptado de HAYKIN (2001)

Conforme Ferneda (2006), o comportamento das conexões entre os neurônios é simulado por meio de seus pesos. Os valores de tais pesos podem ser negativos ou positivos, dependendo de as conexões serem inibitórias ou excitatórias. O efeito de um sinal proveniente de outro neurônio é determinado pela multiplicação do valor (intensidade) do sinal recebido pelo peso da conexão correspondente ( $x_1 \times p_1$ ). É efetuada a soma dos valores  $x_1 \times p_1$  de todas as conexões, e o valor resultante é enviado para a função de ativação, que define a saída ( $y$ ) do neurônio. Combinando diversos neurônios, forma-se uma RNA.

As RNAs são modelos que buscam simular o processamento de informação do cérebro humano. São compostos por unidades de processamentos simples, os neurônios, que se unem por meio de conexões sinápticas (FIGURA 13).



**FIGURA 13 - REPRESENTAÇÃO SIMPLIFICADA DE UMA REDE NEURONAL ARTIFICIAL**

De uma forma simplificada, uma rede neuronal artificial pode ser vista como um grafo onde os nós são os neurônios e as ligações fazem a função das sinapses.

FONTE: Adaptado de HAYKIN (2001)

As RNAs se diferenciam pela sua arquitetura e pela forma como os pesos associados às conexões são ajustados durante o processo de aprendizado. A arquitetura de uma rede neuronal restringe o tipo de problema no qual a rede poderá ser utilizada, e é definida pelo número de camadas (camada única ou múltiplas camadas), pelo número de nós em cada camada, pelo tipo de conexão entre os nós e por sua topologia (HAYKIN, 2001).

Uma das propriedades mais importantes de uma RNA é a capacidade de aprender por intermédio de exemplos e fazer inferências sobre o que aprendeu, melhorando gradativamente o seu desempenho. As redes neuronais utilizam um algoritmo de aprendizagem cuja tarefa é ajustar os pesos de suas conexões (BRAGA, CARVALHO e LUDEMIR, 2000).

Existem duas formas básicas de aprendizado de redes neurais: aprendizado supervisionado e aprendizado não supervisionado. No aprendizado supervisionado, um agente externo (professor) apresenta à rede neuronal alguns conjuntos de padrões de entrada e seus correspondentes padrões de saída. Portanto, é necessário ter um conhecimento prévio do comportamento que se deseja ou se espera da rede. Para cada entrada, o professor indica explicitamente se a resposta calculada é boa ou ruim. A resposta fornecida pela rede neuronal é comparada à resposta esperada. O erro verificado é informado à rede para que sejam feitos ajustes a fim de melhorar suas futuras respostas (FERNEDA, 2006).

Na aprendizagem não supervisionada, ou aprendizado auto-supervisionado, não existe um agente externo para acompanhar o processo de aprendizado. Neste tipo de aprendizagem, somente os padrões de entrada estão disponíveis para a rede neuronal. A rede processa as entradas e, detectando suas regularidades, tenta progressivamente estabelecer representações internas para codificar características e classificá-las automaticamente. Este tipo de aprendizado só é possível quando existe redundância nos dados de entrada, para que se consiga encontrar padrões em tais dados.

#### 3.5.2.1 *Multilayer Perceptron*

A RNA *Multilayer Perceptron* (MLP) é um sistema de neurônios ligados por conexões sinápticas divididos em neurônios de entrada, que recebem estímulos do meio externo, neurônios internos ou ocultos e neurônios de saída, que se comunica com o exterior. A forma de arranjar *Perceptrons* em camadas é denominada *Multilayer Perceptron*. A MLP foi concebida para resolver problemas mais complexos, os quais não poderiam ser resolvidos pelo modelo de neurônio básico. Um único *Perceptron* ou uma combinação das saídas de alguns *Perceptrons* poderia realizar uma operação XOR, porém, seria incapaz de aprendê-la. Para isto são necessárias mais conexões, os quais só existem em uma rede de *Perceptrons* dispostos em camadas (HAYKIN, 2001).

Os neurônios internos são de suma importância na rede neuronal, pois sem estes se torna impossível a resolução de problemas linearmente não separáveis. Em outras palavras, pode-se dizer que uma rede é composta por várias unidades de processamento, cujo funcionamento é bastante simples.

Essas unidades, geralmente são conectadas por canais de comunicação que estão associados a determinado peso. As unidades fazem operações apenas sobre seus dados locais, que são entradas recebidas pelas suas conexões. O comportamento inteligente de uma RNA vem das interações entre as unidades de processamento da rede.

A maioria dos modelos de redes neuronais possui alguma regra de treinamento, onde os pesos de suas conexões são ajustados de acordo com os padrões apresentados. Em outras palavras, elas aprendem através de exemplos. Arquiteturas Neuronais são tipicamente organizadas em camadas, com unidades que podem estar conectadas às unidades da camada posterior.

A rede neuronal passa por um processo de treinamento a partir dos casos reais conhecidos, adquirindo, a partir daí, a sistemática necessária para executar adequadamente o processo desejado dos dados fornecidos. Sendo assim, a rede neuronal é capaz de extrair regras básicas a partir de dados reais, diferindo da computação programada, onde é necessário um conjunto de regras rígidas pré-fixadas e algoritmos (HAYKIN, 2001).

Um algoritmo utilizado na rede neuronal MLP é o de *Levenberg-Marquardt*. Em matemática e computação o algoritmo *Levenberg-Marquardt* (LM) fornece solução para o problema de minimização de uma função, geralmente não linear, mais um espaço de parâmetros da função. Esses problemas surgem principalmente em minimização de mínimos quadrados de ajuste de curva e de programação não linear (LEVENBERG, 1944).

O algoritmo LM alterna entre o algoritmo de Gauss Newton (GN) e o método de gradiente descendente. O algoritmo LM é mais robusto do que o Gauss Newton, o que significa que em muitos casos, ele encontra uma solução mesmo que muito distante do mínimo final. Para funções e parâmetros razoáveis de partida, o algoritmo LM tende a ser um pouco mais lento do que o Gauss Newton. Algoritmo LM também pode ser visto como Gauss Newton usando uma abordagem de intervalo de confiança. O algoritmo LM é usado para resolver problemas genéricos de ajuste de curva. No entanto, o LM encontra apenas um mínimo local, e não um mínimo global (LEVENBERG, 1944).

### 3.5.2.2 Neuro-Fuzzy

Uma rede *Neuro-Fuzzy* pode ser definida como sendo um sistema *Fuzzy* que é treinado por algum algoritmo derivado da teoria das redes neuronais. Este tipo de rede é o resultado da união destes dois modelos. Com isto, tem-se a capacidade das redes neuronais em reconhecimento e classificação, sem esquecer-se da robustez e habilidade de generalização (NAUCK e KRUSE, 1997). No Japão é comum encontrar eletrodomésticos sendo lançados com recursos de autocontrole, por eles chamados *Neuro-Fuzzy*. *Neuro-Fuzzy* são sistemas híbridos combinando redes neuronais com lógica nebulosa ou difusa (PAL e MITRA, 1999).

A rede do tipo *Free Associative Neurons (FAN)* é uma rede neuronal constituído por neurônios independentes com capacidade autônoma de aprendizado que leva em conta a lógica *Fuzzy* (ZADEH, 1994). É uma abordagem *Neuro-Fuzzy* (PAL, 1999) que faz o reconhecimento de padrões.

Em sua tese de doutorado, RAITTZ abordou o uso da rede *FAN* para RP (RAITTZ, 2002). Em 2006 adotando esse modelo de rede neuronal foi desenvolvido um aplicativo na linguagem de programação Java chamado de EasyFAN.

### 3.5.3 Integração em Inteligência Computacional

Várias tentativas foram feitas durante a década de 90 para desenvolver diferentes abordagens híbridas de reconhecimento de padrões, combinando as vantagens de técnicas individuais. Uma integração criteriosa de redes neuronais e teoria *Fuzzy*, comumente conhecida como computação *Neuro-Fuzzy*, é um desses paradigmas híbridos, que é o mais evidente e tem sido adequadamente investigado. Isso permite incorporar as vantagens genéricas de RNAs e uso de lógica *Fuzzy* como paralelismo massivo, robustez, aprendizagem e manuseamento de incerteza e imprecisão, no sistema. Além disso, alguns méritos específicos da aplicação também podem ser incorporados. Por exemplo, no caso de classificação de padrões e geração de regras, pode-se explorar a capacidade de redes neuronais na geração de limites de decisão altamente não-lineares, modelar as incertezas na descrição de entrada e na decisão de saída pelo conceito de conjuntos *Fuzzy* (COX, 1994). Os modelos *Neuro-Fuzzy* são encontrados para obter um melhor desempenho do que uma rede neuronal ou um sistema *Fuzzy* considerados individualmente.



Na realidade existe uma combinação simbiótica de diferentes ferramentas de inteligência computacional para processamento de informação flexível, a fim de lidar com situações ambíguas da vida real e alcançar soluções tratáveis, robustas e de baixo custo (ZADEH, 2001).

A hibridação *Neuro-Fuzzy* é feito em geral de duas maneiras: uma rede neuronal equipada com a capacidade de manipulação da informação *Fuzzy* denominado *Fuzzy Neural Network* (FNN), e um sistema *Fuzzy* acrescido de redes neurais para melhorar algumas características como flexibilidade, velocidade e adaptabilidade; denominado FNS ou *Fuzzy Neural System* (PAL e MITRA, 1999). Em um FNN, quer os sinais de entrada e / ou pesos de conexão e / ou saídas são subconjuntos *Fuzzy* ou valores de pertinência a alguns conjuntos *Fuzzy* (PAL e SKOWRON, 1999). Normalmente, os valores linguísticos como: baixa, média e alta, ou números *Fuzzy* ou intervalos são usados nesta modelagem. Redes neurais com neurônios *Fuzzy* são também chamados FNN uma vez que eles são capazes de processar informação *Fuzzy*. Um FNS, por outro lado, é projetado para realizar o processo de raciocínio *Fuzzy*, onde os pesos de conexão da rede correspondem aos parâmetros do raciocínio *Fuzzy*. A arquitetura do FNS possui nós distintos para cláusulas antecedentes, operadores de conjunção e cláusulas de consequência.

Posteriormente, a hibridização dos conjuntos *Fuzzy* tem sido investigada utilizando outras ferramentas de inteligência computacional, como Algoritmos Genéticos (PAL e SKOWRON, 1999).

### 3.6 TRABALHOS APROXIMADOS COM REDES NEURONAIS ARTIFICIAIS

Foram encontrados estudos com uso de redes neurais artificiais como apoio a decisão no diagnóstico. Alguns trabalhos aproximados e interessantes que podem se referenciar é: Diagnóstico Mamográfico (AGOSTI, BAYS e MORENO, 2004), Diagnóstico Diferencial da Doença Meningocócica (MONAT, MARTINS e ASSIS, 2004), Diagnóstico Diferencial de Lesões Intersticiais Pulmonares (AMBRÓSIO, 2002), Diagnóstico de Doenças Reumatológicas (FALQUETO, SOUZA e TOSTES, 2003). Não foi encontrado, especificamente, o uso de redes neurais para apoio a decisão na liberação de exames laboratoriais clínicos do perfil lipídico e glicemia em jejum.

Além dos trabalhos citados anteriormente, foram encontrados algumas ferramentas para detecção de cardiopatias que tinham por objetivo aperfeiçoar o diagnóstico médico com o uso de redes neuronais. Dentre as ferramentas que serão citadas, além de outras doenças, geralmente todas estão focadas na análise de Eletrocardiograma (ECG), são elas: Solução de Soares e Nadal, a ferramenta de Sivathanan, Cecelja e Balachandran, o WaveECG e por fim o DoctorCardio. Para melhor visualizar segue abaixo a Tabela 4 com os comparativos dos trabalhos, sendo 2 deles relativos à doença de Infarto Agudo do Miocárdio (IAM).

**TABELA 4 - ESTUDOS COM REDES NEURONAIS PARA IDENTIFICAÇÃO DE PADRÕES ASSOCIADOS AO DIGNÓSTICO CLÍNICO**

<b>Trabalho</b>	<b>Aplicação</b>	<b>Treinamento</b>	<b>Acurácia</b>
Soares e Nadal (1999)	Doenças relacionadas à alteração do segmento ST.	<i>Levenberg-Marquardt (LM)</i>	88,7%
Sivathanan, Cecelja e Balachandran (2000)	Cinco doenças cardíacas.	<i>Backpropagation</i>	Não Divulgado
WaveECG (2008)	IAM	-	88,5%
DoctorCardio (2008)	IAM	<i>Backpropagation</i>	91,0%

IAM, Infarto Agudo do Miocárdio.

FONTE: O autor (2011)

Conforme Tabela 4, para a solução de Soares e Nadal (1999) foi utilizado o algoritmo *Levenberg-Marquardt (LM)*. Com esse método a solução conseguiu atingir um taxa de acerto global (acurácia) de 88,7%.

Seguindo a mesma linha, e utilizando algoritmos inteligentes para diagnóstico de doenças cardíacas, Sivathanan, Cecelja e Balachandran (2000) utilizaram RNAs, porém para decisão final do diagnóstico foi utilizada lógica *Fuzzy*. O treinamento das RNAs foi realizada na rede MLP com algoritmo *Backpropagation*.

As doenças, derivações analisadas e taxa global de acerto não foram divulgados (SIVATHASAN, CECELJA e BALACHANDRAN, 2000). O *software* WaveECG desenvolvido por Dourado, Souza e Oliveira, teve a taxa de acerto global (acurácia) de 88,5% (DOURADO, SOUZA e OLIVEIRA, 2008).

Por fim o DoctorCardio, que foi desenvolvido por Rosário, Monteiro e Gomes, utilizando para treinamento o algoritmo de *Backpropagation*, teve taxa global de acerto de 91,0%, (ROSÁRIO, MONTEIRO e GOMES, 2008).

Como trabalho similar pode-se citar a dissertação de Volanski que utilizou as redes neuronais MLP e FAN para avaliar ensaios de Urinálise (VOLANSKI, 2011). Volanski (2011) usou como critérios de grupos ou classes, os resultados dos profissionais designados como exames “liberados” ou “retidos”. As redes neuronais estudadas apresentaram acurácia de 90,9% e 82,8%, respectivamente para as redes MLP e FAN.

O desempenho das redes neuronais estudadas por Volanski (2011) foram comparadas e avaliadas em uma amostra de 484 pacientes da rotina por profissionais. A rede MLP apresentou uma divergência de cerca de 9% a 12% (sensibilidade 100%; especificidade 90,1%) e com os profissionais, significativamente menor que a rede FAN, 17% a 19% (sensibilidade 39,0% e especificidade 86,9%), respectivamente.

A comparação entre o desempenho das redes neuronais com o banco de dados, superior a 100 mil registros, mostrou que a rede MLP apresentou desempenho compatível ao dos profissionais (29,5% vs. 21,6%, respectivamente) na classificação dos ensaios retidos (VOLANSKI, 2011).

A rede FAN reteve menos amostras (18,2%) sugerindo a maior divergência com a resposta dos profissionais. A rede neuronal MLP validada com uma base de dados “balanceada” permite identificar padrões “liberação” e “retenção” de resultados de forma a mimetizar os critérios de profissionais para o ensaio de Urinálise (VOLANSKI, 2011).

## 4 MATERIAIS E MÉTODOS

O projeto dissertado tem a aprovação do Comitê de Ética em Pesquisa com Seres Humanos do Hospital de Clínicas da Universidade Federal do Paraná sob número CAE: 0253.0.208.000-10 (ANEXO).

### 4.1 AMOSTRA

Foram utilizados três conjuntos de registros de resultados de exames do setor de Imunoquímica do Laboratório do Hospital de Clínicas da Universidade Federal do Paraná armazenados no banco de dados DB2 (IBM) hospedado em Servidor AIX (IBM), a partir de uma amostra total de 357.893 registros (linhas). Foi considerado jejum prévio para glicemia de 8 horas no mínimo e de 12 a 14 horas para o perfil lipídico respeitando a IV Diretrizes Brasileiras em Dislipidemia (2007).

Todos os conjuntos continham as informações sobre idade e sexo.

Todos os resultados das concentrações de glicemia em jejum, colesterol total, HDL-colesterol (HDL-c), LDL-colesterol (LDL-c) e triglicérides (TG) estavam presentes nos arquivos denominados “completo”.

Os arquivos denominados como “incompleto” foram caracterizados por apresentarem as informações sobre idade e sexo e não necessariamente todas as informações sobre os elementos que compõem o perfil lipídico e glicemia em jejum. No arquivo incompleto, pelo menos um resultado dos elementos em estudo estava presente. Este arquivo representa a rotina hospitalar aonde nem sempre todos os exames são solicitados em conjunto pelo clínico.

Todos os dados extraídos e analisados não disponibilizam os nomes dos pacientes, preservando o anonimato destes. Os conjuntos de arquivos foram separados em “completo” e “incompleto” com as características sumarizadas na Figura 14.

Os dados extraídos do banco de dados DB2 do HC-UFPR são relativos ao período de 2006 a 2010. O estudo com arquivos completos e incompletos visa verificar o desempenho das redes neuronais com todos os parâmetros ou concentrações informados e também com parâmetros incompletos para quantidades maiores de registros, mimetizando a rotina laboratorial.

**Conjunto completo:** O conjunto principal composto por 60.006 registros que contempla os resultados dos pacientes que apresentaram todos os parâmetros bioquímicos em estudo, supracitados, em uma mesma solicitação clínica. Este conjunto foi utilizado para treinamento e teste das redes neurais em estudo.

**Conjuntos incompletos:** Outros dois conjuntos foram obtidos do banco de dados do HC-UFPR e contemplam um tamanho amostral de 131.072 e 240.000 registros. Estes conjuntos de dados, não estão completos de todos os parâmetros e representam as frequências ou as rotinas com que os exames estudados neste trabalho são liberados no setor de Imunoquímica do HC-UFPR.

Para obter os registros, os ensaios foram selecionados do banco de dados DB2 do HC-UFPR através de comandos SQL (APÊNDICE B) e exportado para arquivo do *software* Microsoft® Excel, onde foram processados posteriormente.

Na planilha Excel, usando recursos de Tabela Dinâmica, foi gerado um conjunto com 60.006 resultados de exames, utilizado no treinamento ( $n=30.003$ ) e teste ( $n=30.003$ ) das redes neurais. O arquivo foi importado para uma base de dados previamente modelada no MYSQL que foi designado de “GLICEMICO\_LIPIDICO” tendo como tabela com os parâmetros em estudo designada “EXAMES”. O *script* para a geração da tabela “EXAMES” está descrito no Apêndice B. Uma vez os dados importados no MYSQL, foram trabalhados usando comandos de SQL (APÊNDICE B) para obter a classificação.

Para os arquivos incompletos, foram selecionados sequencialmente 131.072 e 240.000 registros. O objetivo da amostra é constatar como se comporta a rede neuronal com quantidades maiores de registros ao serem submetidas ao treino e teste nas redes neurais.

Foram selecionados 357.893 registros, destes foram utilizados 240.000 registros para os conjuntos incompletos, ou seja, dentro do universo dos 240.000 registros contemplavam os 131.072 registros usados para os arquivos I-1 e I-2 conforme Figura 14.

Para os arquivos completos, foram selecionados 60.006 registros que contemplavam todos os parâmetros em estudo preenchidos, que foram retirados do conjunto de 357.893 registros. O arquivo completo de 60.006 registros foi subdividido em dois arquivos iguais de 30.003 registros para treinamento e teste (C-1 e C-2, FIGURA 14).

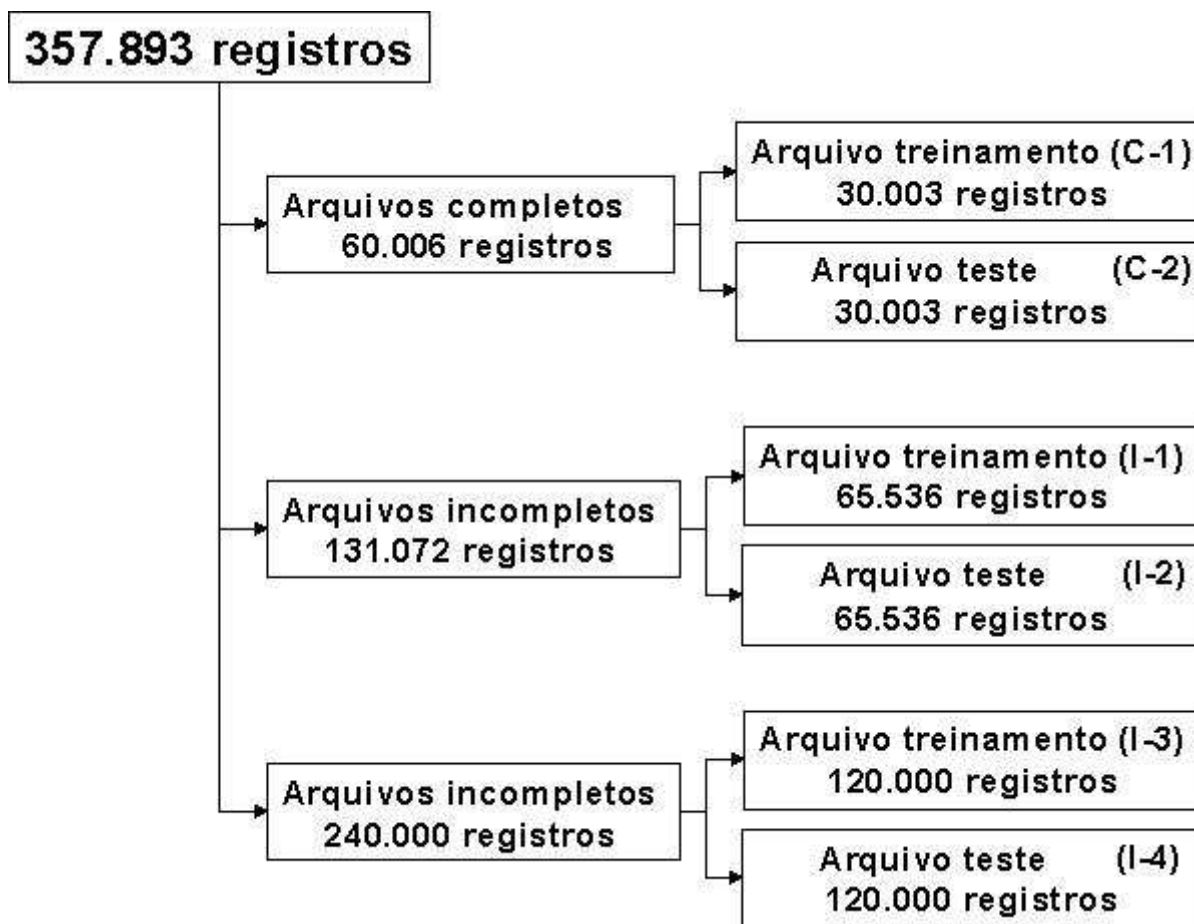


FIGURA 14 - CLASSIFICAÇÃO DOS ARQUIVOS DE DADOS UTILIZADOS PARA ESTUDOS COM REDES NEURONAIS.

A partir do banco de dados do HC-UFPR foram selecionados sequencialmente 357.893 registros para os parâmetros em estudo, todos contendo as informações de sexo e idade. Estes registros foram classificados em três arquivos, um completo (C) e dois incompletos (I).

O arquivo completo, contendo todos os parâmetros estudados para cada linha do banco foi subdividido em arquivos de treino (C-1) e teste (C-2).

Os arquivos incompletos, contem em cada linha, valores de resultados variáveis para os exames, podendo ser 1 a 5. Todos os arquivos possuíam as características de sexo e idade. Estes foram subdivididos em I-1 a I-4.

FONTE: O autor (2011)

Para as amostras treinadas e testadas (arquivos completos e incompletos), não foram utilizados, respectivamente, todos os registros selecionados no período de 2006 a 2010.

## 4.2 CLASSIFICAÇÃO DA AMOSTRA

Os registros no banco de dados foram classificados com os critérios descritos da Tabela 5. Os valores apresentados foram obtidos de consensos internacionais para caracterização de alterações patológicas (valores de corte superior) ou valores empíricos de uso comum entre profissionais, especificamente para os valores de concentração baixa no perfil lipídico.

Pacientes que apresentam todos os resultados dentro dos critérios estabelecidos são liberados (Liberados; código=0) e aqueles que estão fora dos parâmetros estabelecidos, mesmo que para um único analito são classificados como retidos (Retidos, código=1). Os resultados retidos são reavaliados pelos profissionais.

TABELA 5 - CRITÉRIOS PARA A CLASSIFICAÇÃO DA AMOSTRA

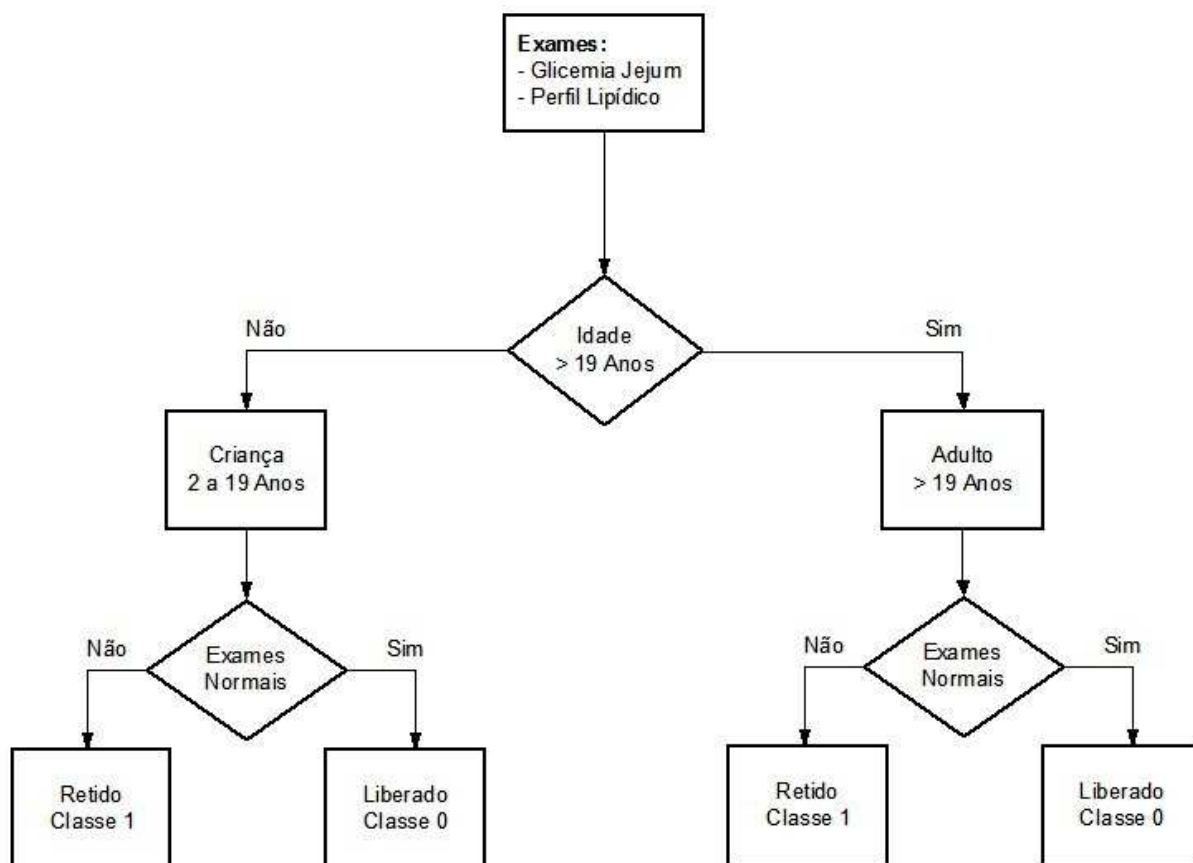
<b>ANALITOS</b>	<b>ADULTOS &gt;19 ANOS</b>	<b>CRIANÇAS 2 a 19 ANOS</b>
Glicemia em Jejum (mg/dL) *	50 – 126	50 – 100
Colesterol Total (mg/dL) **	100 – 240	60 – 170
HDL-Colesterol (mg/dL) <sup>a</sup>	15 – 100	15 – 100
LDL-Colesterol (mg/dL) **	25 – 130	25 – 110
Triglicérides (mg/dL) **	10 – 150	10 – 100

\* Para glicemia foram utilizados os critérios de diagnóstico de diabetes (>126 mg/dL) e hipoglicemia (<50 mg/dL)

\*\*Para o perfil lipídico os limites superiores foram baseados nos valores descritos na IV Diretrizes Brasileiras em Dislipidemia (2007) e os valores inferiores são empíricos, de uso corrente entre profissionais do laboratório.

<sup>a</sup> Os valores de corte para HDL-colesterol foram empíricos, com base na experiência profissional.  
FONTE: O autor (2011)

A partir dos dados coletados, os mesmos foram submetidos à classificação conforme critérios dos resultados dos exames de glicemia e perfil lipídico. Com base nos critérios acima os exames foram classificados respeitando o fluxograma conforme Figura 15.



**FIGURA 15 - CLASSIFICAÇÃO DOS EXAMES PARA REDE NEURONAL**

Os dados de glicemia em jejum e do perfil lipídico foram inicialmente classificados por idade (>19 anos) e posteriormente utilizando os valores de referência para os parâmetros em estudo. Valores fora da referência são classificados como retidos (classe=1) e resultados dentro da referência como liberados (classe=0).

FONTE: O autor (2011)

Após a classificação, a base ficou composta de linhas (registros), com os dados dos exames de perfil lipídico e glicemia. Estas linhas contiveram colunas com as seguintes informações:

1. Sexo do paciente;
2. Idade do paciente;
3. Resultado do exame de Glicemia em Jejum;
4. Resultado do exame de HDL-colesterol;
5. Resultado do exame de LDL-colesterol;
6. Resultado do exame de Triglicérides;
7. Resultado do exame de Colesterol Total;
8. Classificação dos exames entre Liberado e Retido.



Portanto, foram utilizadas 7 características mais a classe conforme Tabela 6. Inicialmente esta base de dados foi dividida igualmente em duas partes, formando a base de treinamento e a base de teste. Então se submeteu à base de treinamento nas duas redes neurais: MLP e FAN.

**TABELA 6 - EXEMPLO DA DISTRIBUIÇÃO DOS DADOS USADOS PARA TREINAMENTO DA REDE PARA ARQUIVO COMPLETO (C-1)**

<b>Sexo</b>	<b>Idade</b>	<b>Glicemia</b>	<b>HDL-c</b>	<b>LDL-c</b>	<b>TG</b>	<b>CT</b>	<b>Classe</b>
1	47	54	94	10	56	115	1
2	13	97	44	76	101	140	1
1	15	92	39	128	124	192	1
1	16	98	43	105	90	166	0
1	47	104	165	44	149	239	1
2	12	100	50	97	53	158	0
2	68	49	51	48	31	105	1
2	20	91	34	72	150	136	0
1	56	52	57	114	47	180	0
1	12	100	57	71	95	147	0
1	16	73	50	128	41	186	1

A primeira característica (coluna) foi convertida de letra para número. A rede neuronal trabalha melhor com número do que letras. O número 1 corresponde ao sexo Feminino e o número 2 corresponde ao sexo Masculino. A última coluna corresponde à classe onde o número 0 (zero) corresponde a “liberado” e o número 1 a “retido”.

HDL-c, HDL-colesterol; LDL-c, LDL-colesterol; TG, triglicérides; CT, colesterol total.

FONTE: O autor (2011)

A Tabela 7 mostra exemplo de arquivo incompleto que foi utilizado para comparação de desempenho nas redes neurais MLP e FAN em relação ao arquivo completo.

TABELA 7 - EXEMPLO DA DISTRIBUIÇÃO DOS DADOS USADOS PARA TREINAMENTO DA REDE PARA ARQUIVO INCOMPLETO (I-1)

<b>Sexo</b>	<b>Idade</b>	<b>Glicemia</b>	<b>HDL-c</b>	<b>LDL-c</b>	<b>TG</b>	<b>COL</b>	<b>Classe</b>
1	78	105	0	0	0	0	0
1	78	2004	0	0	0	0	1
2	66	98	0	0	0	0	0
1	51	0	29	0	657	201	1
1	52	0	-9999	0	780	276	1
1	54	224	0	0	2482	449	1
1	54	182	0	0	575	292	1
2	46	0	44	59	226	148	1
2	46	0	0	0	1520	310	1
1	55	0	36	0	0	226	1
1	43	0	46	147	122	217	1

Pelo termo “Arquivo Incompleto” entenda-se, em comparação com a Tabela 6, que valores para alguns exames estão zerados, possuindo ainda valores negativos e valores muito acima da realidade. Os valores zerados foram preenchidos pela ausência de resultados de exames. A primeira característica (coluna) foi convertida de letra para número. O número 1 corresponde ao sexo Feminino e o número 2 corresponde ao sexo Masculino. A última coluna corresponde à classe onde o número 0 (zero) significa “liberado” e o número 1 “retido”.

HDL-c, HDL-colesterol; LDL-c, LDL-colesterol; TG, triglicérides; COL, colesterol total.

FONTE: O autor (2011)

As bases de dados foram treinadas e testadas com os arquivos classificados como descrito anteriormente. Estes arquivos foram submetidos às redes neurais com os *softwares* EasyFAN (rede neuronal FAN) e MATLAB (rede neuronal MLP).

#### 4.3 DOSAGENS LABORATORIAIS

As determinações dos parâmetros laboratoriais avaliados neste estudo glicemia em jejum, colesterol total e triglicérides (TG) foram quantificados por métodos enzimáticos colorimétricos, o HDL-colesterol pelo método homogêneo no sistema automatizado Architect ci8200 da Abbott®, utilizando reagentes, calibradores e controles fornecidos pelo fabricante do equipamento.

A concentração de LDL-colesterol foi calculado utilizando a equação de Friedewald ( $\text{LDL-c} = \text{colesterol total} - \text{HDL-c} - \text{TG}/5$ ), observando as limitações como níveis de Triglicérides  $> 400 \text{ mg/dL}$  conforme preconizado nas IV Diretrizes Brasileiras em Dislipidemia (2007).

## 4.4 EQUIPAMENTOS E SOFTWARES UTILIZADOS

### 4.4.1 Hardware

Para desenvolver o projeto foi usado um Notebook da marca DELL modelo VOSTRO 1310 com processador INTEL CORE 2 DUO com 2.0 GHZ e memória RAM de 4GB DDR2 667MHZ com HD de 250 GB SATA que foi utilizado pelo pesquisador para o tratamento dos dados, treinamento e teste das redes neurais.

### 4.4.2 Softwares

Foram utilizados os *softwares*: SIH (*software* do HC para requisição de exames laboratoriais clínicos); DB2®, SQL, EXCEL®, WAMPSERVER, MYSQL, STATISTICA®, OPENEPI, EASYFAN e MATLAB®.

#### 4.4.2.1 DB2®

Produzido pela IBM®, o DB2 é o Sistema Gerenciador de Banco de Dados (SGBD) relacional. Existem diferentes versões do DB2 que rodam desde num simples computador de mão ou PDA, até em potentes *mainframes* e funcionam em servidores baseados em sistemas operacionais Unix, Windows ou Linux. O nome DB2 foi dado para o Sistema Gerenciador de Banco de Dados que a IBM lançou em 1983, baseado em SQL/DS para seu *mainframe* (IBM, 2010).

O DB2 é oferecido em diversos tipos de edições ou licenças. Pela escolha de uma versão com menos recursos, a IBM evita que os consumidores invistam por recursos que não iriam usar. Alguns exemplos de edição são a *Express*, *Workgroup* e a edição *Enterprise*. A edição mais sofisticada para Linux/UNIX/Windows é o DB2 *Data Warehouse Enterprise Edition* (DB2 DWE). A versão DB2 *Express C* é uma versão gratuita para Linux e Windows podendo baixá-la mediante cadastro (IBM, 2010).

No DB2 estavam armazenados todos os resultados de exames usados na pesquisa da dissertação. O DB2 foi utilizado para extrair os valores de resultados dos exames laboratoriais clínicos através de comandos SQL (APÊNDICE B).

#### 4.4.2.2 SQL

*Structured Query Language* (SQL) é uma linguagem de pesquisa declarativa para banco de dados relacional. Muitas das características originais do SQL foram inspiradas na álgebra relacional.

O SQL foi desenvolvido originalmente no início dos anos 70 nos laboratórios da IBM em San Jose, dentro do projeto *System R*, que tinha por objetivo demonstrar a viabilidade da implementação do modelo relacional proposto por E. F. Codd. O nome original da linguagem era SEQUEL, acrônimo para *Structured English Query Language* (CHAMBERLIN *et al.*, 1981).

A linguagem SQL é padrão de banco de dados. Decorre da sua simplicidade e facilidade de uso. Ela se diferencia de outras linguagens de consulta a banco de dados no sentido em que uma consulta SQL especifica a forma do resultado e não o caminho para chegar a ele. Ela é uma linguagem declarativa em oposição a outras linguagens procedurais. Isto reduz o ciclo de aprendizado daqueles que iniciam na linguagem.

#### 4.4.2.3 EXCEL®

O Microsoft® Excel (<http://www.microsoft.com>) é um *software* de planilha eletrônica de cálculo escrito e produzido para computadores que utilizam o Sistema Operacional (SO) Microsoft® Windows e também computadores Macintosh da Apple®. Seus recursos incluem uma interface intuitiva, capacitadas ferramentas de cálculo e de construção de gráficos que, juntamente com marketing agressivo, tornaram o Excel um dos mais populares aplicativos de computador. É, com grande vantagem, o aplicativo de planilha eletrônica dominante, disponível para essas plataformas, e o tem sido desde a versão 5 em 1993 e sua inclusão como parte do Microsoft® Office.

O Excel foi utilizado na pesquisa da dissertação após a extração dos dados dos exames laboratoriais do banco de dados DB2 e trabalhado na tabulação dos resultados de exames, ou seja, chamado de pareamento o posicionamento do exame ao lado do outro através do recurso de Tabela Dinâmica.

#### 4.4.2.4 WAMPSERVER

É um ambiente de desenvolvimento *web* para uso no Windows. Ele permite a você criar aplicações *web* com Apache, PHP e banco de dados MySQL. Vem disponibilizado com o phpMyAdmin para gerenciar facilmente as suas bases de dados. A instalação do WAMPSERVER e seu uso é muito intuitivo. Geralmente não há necessidade de configuração dos seus arquivos (WAMPSERVER, 2010).

WAMPSERVER é uma solução empacotada que lhe permitirá reproduzir o seu servidor de produção. Depois que o WAMPSERVER é instalado, você terá a possibilidade de acrescentar muitos projetos ou lançamentos como o Apache, MySQL e PHP. Conhecido anteriormente como WAMP5, é publicado sob a licença *General Public License* (GPL) ou Licença Geral Pública. Existem vários pacotes de extensão para realizar download no site oficial, como por exemplo, para ativar o uso de PHP4 ao invés do PHP5. A partir da versão 2, ocorreu a mudança do nome de WAMP5 para WAMPSERVER.

Foi utilizado no projeto juntamente com o MySQL para a classificação dos exames laboratoriais em estudo.

#### 4.4.2.5 MYSQL

É um Sistema Gerenciador de Banco de Dados (SGBD) relacional que utiliza a linguagem padrão SQL e é largamente utilizado em aplicações internet. O MySQL é considerado um banco de dados relacional rápido, que supera quase todos os principais bancos de dados em todas as categorias. É aplicado principalmente em sistemas baseados em UNIX, embora haja portas para quase toda plataforma no mercado (MASLAKOWSKI, 2000).

O MySQL usa a Licença Pública Geral GNU. GNU é um acrônimo de *GNU Not Unix*. A *Free Software Foundation* (FSF) é a principal organização patrocinadora do Projeto GNU. A FSF recebe poucos fundos de grandes empresas ou organizações. Ela depende do suporte de pessoas voluntárias para cumprir a missão de preservar, proteger e promover a liberdade de usar, estudar, copiar, modificar, redistribuir *software* para computadores e para defender os direitos dos usuários de *software* livre.

As principais características que tornam o MySQL tão popular são:

- Facilidade de instalação, uso e administração;
- Desempenho mesmo com poucos recursos de máquina;
- Robustez, apresentando também alta escalabilidade, sendo capaz de lidar com grandes volumes de dados sem que haja o comprometimento da integridade, nem do seu desempenho para a manipulação desses dados;
- Flexibilidade de uso em diversas plataformas;
- Flexibilidade em permitir tipos de tabelas diferentes (MyISAM, InnODB, Cluster, MaxDB) para diferentes aplicações.

Para o projeto de pesquisa foi utilizado o MySQL juntamente com WAMP SERVER para o armazenamento dos dados e a classificação dos exames em estudo.

#### 4.4.2.6 STATISTICA®

É uma ferramenta abrangente utilizada para análise integrada de dados, gráficos, histogramas, gerenciamento de dados e probabilidades. É um *software* com uma enorme variedade de procedimentos básicos e avançados de análise de negócio, entre elas, de mineração de dados, ciência e aplicações de engenharia (STATSOFT, 2011).

A estatística do projeto de pesquisa foi realizada utilizando este *software*, conforme exemplo nas Tabelas 11 e 12 constadas no item 5 (RESULTADOS). A versão do *software* utilizada na pesquisa é a 8.0.

#### 4.4.2.7 OPENEPI

O *software Open Source Epidemiologic Statistics for Public Health* fornece suporte estatístico para a contagem e medições em estudos analíticos, tamanho da amostra e realiza cálculos com números aleatórios tais como sensibilidade, especificidade, qui-quadrado e outras análises estatísticas (DEAN, SULLIVAN e SOE, 2011).

OpenEpi é um *software* livre para as estatísticas epidemiológicas. Pode ser executado em um servidor web ou baixado e executado sem conexão *web*. Os programas são escritos em *JavaScript* e HTML, sendo compatível com Linux, Mac e navegadores do PC, independentemente do sistema operacional (DEAN, SULLIVAN e SOE, 2011).

Os programas têm uma licença de código aberto e pode ser baixado, distribuído, ou modificado. Alguns dos componentes de outras fontes têm licenciamento declarado nos arquivos de código-fonte. O *software* OPENEPI foi desenvolvido com apoio em parte da doação da Fundação Bill e Melinda Gates para a Universidade de Emory, Atlanta (DEAN, SULLIVAN e SOE, 2011).

O *software* foi utilizado para calcular os valores de sensibilidade, especificidade, valor preditivo positivo e negativo bem como a acurácia das redes neurais artificiais. A versão do *software* utilizada foi a 2.3.1.

#### 4.4.2.8 MATLAB®

É um *software* destinado a fazer cálculos com matrizes (acrônimo: MATrix LABoratory). É uma linguagem de alto nível e um ambiente interativo que se propõe a executar tarefas de computação intensiva mais rapidamente que as linguagens de programação tradicionais como C, C++ e Fortran (MATHWORKS, 2010).

Inúmeras aplicações se podem utilizar o MATLAB, entre os quais: processamento de imagem, construção de gráficos, modelagem, análise financeira e biologia computacional. Possui bibliotecas (*toolboxes*) que são coleções de funções especiais codificadas em MATLAB. Além disso, é possível inserir novas funções e estender ao ambiente de desenvolvimento para resolver determinadas classes de problemas na área de aplicação desejada.

MATLAB fornece um número de características para documentar e partilhar o trabalho. Pode integrar o código MATLAB com outras linguagens e aplicativos, e distribuir os seus algoritmos e aplicações (MATHWORKS, 2010).

O MATLAB possui como principais características as seguintes:

- Linguagem de alto nível para computação técnica;
- Ambiente para o gerenciamento de código, arquivos e dados;
- Ferramentas interativas para exploração, design e resolução de problemas;
- Funções matemáticas para álgebra linear, estatísticas, análise de Fourier, filtragem, otimização e integração numérica;
- Gráficos 2-D e 3-D para visualização de dados;
- Ferramentas para a construção de interfaces gráficas personalizadas;
- Funções de integração MATLAB baseado em algoritmos com aplicações externas e linguagens, como C, C ++, Fortran, Java, COM e Microsoft Excel.

A Figura 16 apresenta o ambiente de trabalho do MATLAB.

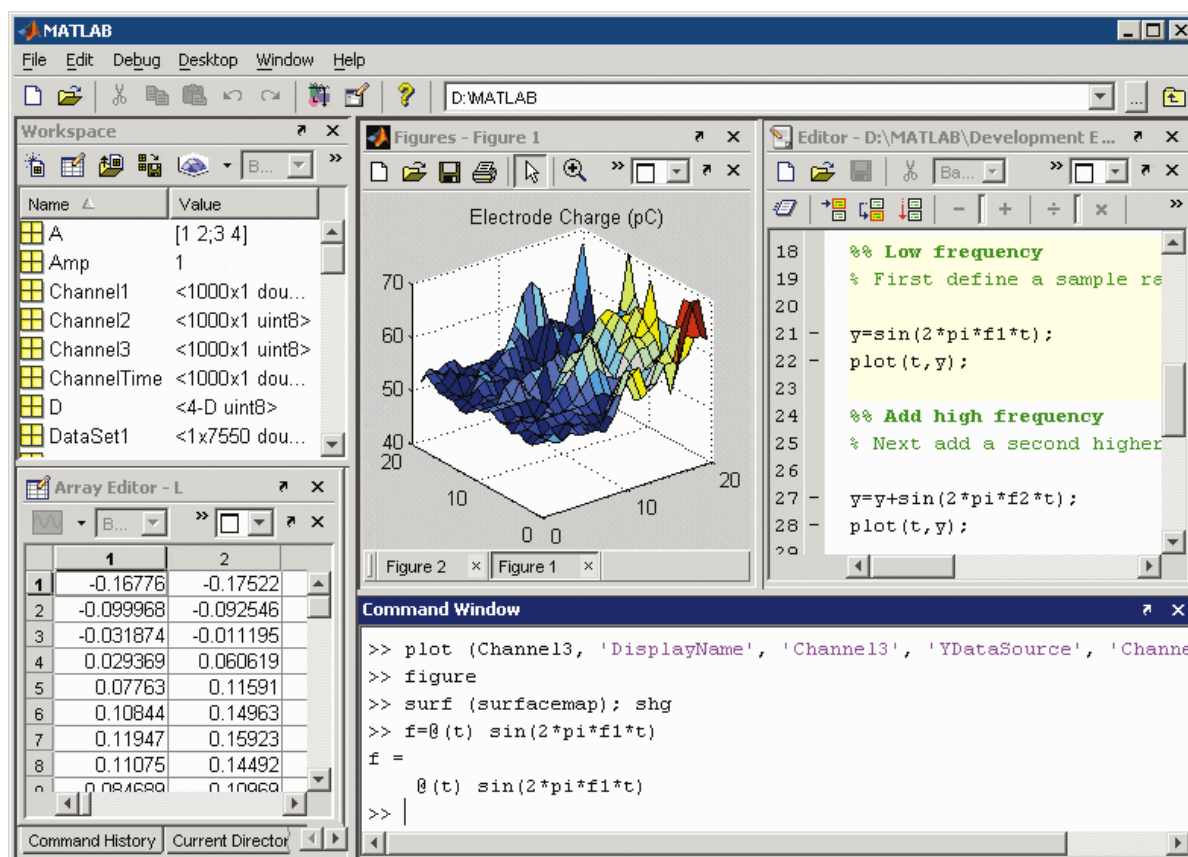


FIGURA 16 - AMBIENTE DE DESENVOLVIMENTO DO SOFTWARE MATLAB®  
 FONTE: Livre adaptação de <<http://www.mathworks.com/products/matlab/>>



O *software* MATLAB ® R2010a foi utilizado usando *scripts* ou comandos para treinar e testar rede neuronal do tipo MLP. A extração das divergências da rede neuronal MLP foi obtida através de *scripts* no MATLAB (APÊNDICE C). Posteriormente foi comparado seu desempenho com a rede neuronal FAN.

#### 4.4.2.9 EASYFAN

É um projeto que tem por base o reconhecimento de padrões sem que o usuário saiba as definições básicas de inteligência artificial, o que é reconhecimento de padrões, tendo facilidade na abertura de arquivos e navegação entre as telas (KUSTER *et al.*, 2006). Foi concebido o projeto de *software* em 2006 utilizando a linguagem de programação Java (FIGURA 17).

FAN é uma abordagem *Neuro-Fuzzy* que faz o reconhecimento de padrões. Seu objetivo principal é a classificação. A abordagem comum em reconhecimento de padrões é dividir o problema em duas partes: extração das características e classificação (RAITZ, 1997). Esta rede foi utilizada como mecanismo para reconhecimento de padrões entre a classificação dos valores de referência dos exames liberados e exames com valores fora da faixa normal ou retidos.

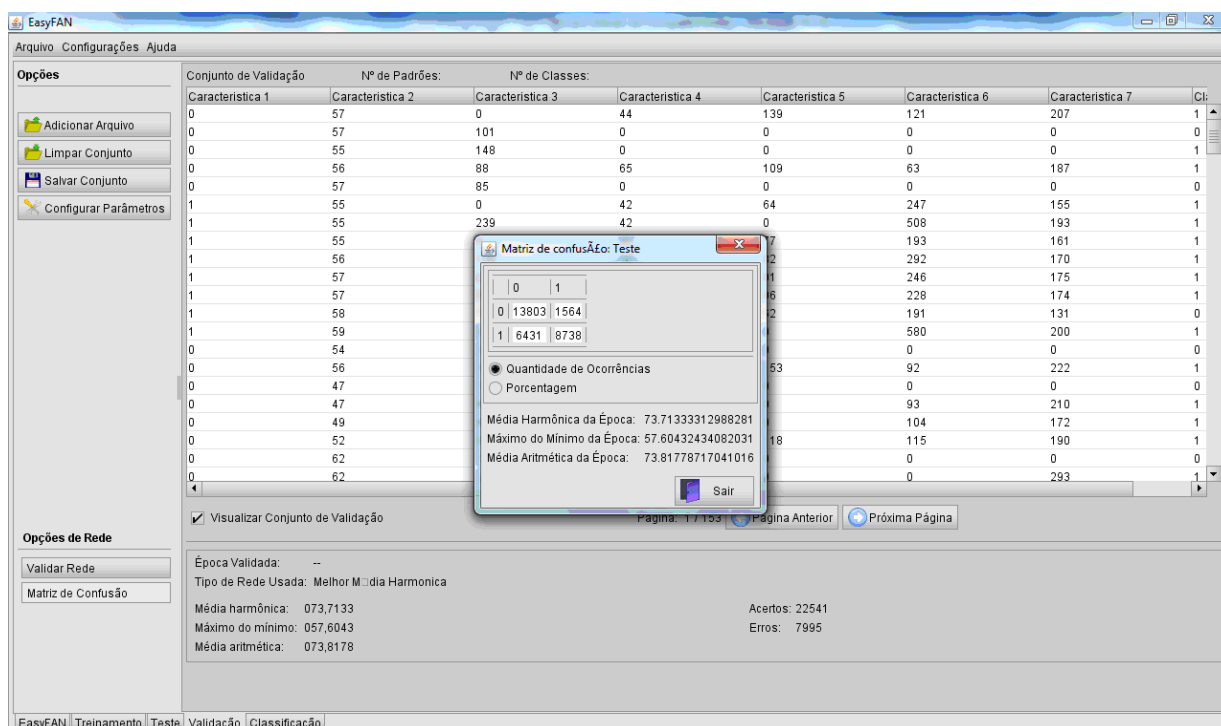


FIGURA 17 - DEMONSTRATIVO DA TELA DO SOFTWARE EASYFAN

Os dados foram transportados para a Tabela 15 (RESULTADOS).

O EasyFAN pode ser encontrado em <<http://sourceforge.net/projects/easyfan>>.

FONTE: KUSTER e colaboradores (KUSTER *et al.*, 2006)

O *software* EasyFAN foi utilizado para treinamento e teste da rede neuronal do tipo FAN. O resultado encontrado na rede neuronal FAN através do *software* EasyFAN foi comparado com a rede neuronal MLP. O *software* pode ser obtido ainda no endereço: [http://www.bioinfo.ufpr.br/sibila/arq/easyFanFreeze5\\_6\\_2.rar](http://www.bioinfo.ufpr.br/sibila/arq/easyFanFreeze5_6_2.rar)

#### 4.4.3 Treinamento e Teste das Redes Neurais

Foram selecionados 60.006 exames de perfil lipídico e glicemia em jejum. Estudando a base completa, verificou-se a proporção de 22.719 registros com exames classificados como liberados, representando 37,2% da amostra e foram encontrados 37.287 registros correspondendo a 62,8% da amostra classificada como retidos. Durante o treinamento da rede neuronal utilizou-se as bases de treinamento e teste com 30.003 registros cada.

Na rede MLP, foram utilizadas várias formas de montagem, como por exemplo, mudando de posição as características. Nesse estudo utilizou-se 3 camadas ocultas: 1 camada oculta de 7 *perceptrons* e duas camadas ocultas com 15 *perceptrons* cada para 500 épocas ou ciclos. Um número excessivo de ciclos pode levar a rede à perda do poder de generalização (*overfitting*). Por outro lado, com um pequeno número de ciclos a rede pode não chegar ao seu melhor desempenho (*underfitting*). Sugere-se um valor entre 500 e 3.000 ciclos de treinamento (SILVA e OLIVEIRA, 2011). O algoritmo utilizado para rede neuronal MLP foi *Levenberg-Marquardt*. *Scripts* foram criados no MATLAB (APÊNDICE C).

Na rede do tipo FAN, os mesmos arquivos foram carregados no aplicativo EasyFAN e executado o treinamento com 500 épocas ou ciclos. Foi utilizado a média harmônica, por ter o melhor desempenho, para testar os dados processados no EasyFAN.

#### 4.5 DIVERGÊNCIAS DA REDE MLP ANALISADAS POR PROFISSIONAIS ESPECIALISTAS

Dois profissionais especialistas do Setor de Imunoquímica do HC-UFPR analisaram registros que apresentaram divergência (n=377) na rede MLP para a classificação através dos valores de referência.

A matriz de divergência (matriz de confusão) foi construída através de *scripts* realizados no MATLAB (APÊNDICE C) e os resultados são exibidos na Tabela 13.

Os profissionais receberam o arquivo enviado pelo pesquisador, aonde o critério de liberação foi omitido e foram solicitados a verificar o motivo da divergência ocorrida pela rede MLP, aquela que apresentou a menor taxa de divergência. Foram apresentados para os profissionais os valores de corte utilizado (TABELA 5) para a classificação dos exames como “liberados” e “retidos”. Baseados nestes valores e na experiência de cada profissional, os mesmos enviaram o arquivo para o pesquisador com sugestões ou observações e motivos da divergência da rede neuronal MLP.

Os resultados enviados pelos profissionais especialistas são mostrados e explicados no item 5.3 (RESULTADOS).

#### 4.6 ANÁLISES ESTATÍSTICAS

Os valores que compõem a tabela 2 x 2 para os cálculos de sensibilidade e outros indicadores com o OPENEPI (<http://www.openepi.com>) foram obtidos através de Matriz de Confusão (TABELA 8). Os valores mostrados na Figura 18 refletem a realidade da Tabela 13 para a rede neuronal MLP.

TABELA 8 - ESTRUTURA DA MATRIZ DE DIVERGÊNCIA UTILIZADA NO PROJETO

Resultados	Rede Neuronal (FAN e MLP)	
	Positivo	Negativo
Positivo	<b>a</b> Verdadeiro Positivo	<b>b</b> Falso Positivo
Negativo	<b>c</b> Falso Negativo	<b>d</b> Verdadeiro Negativo

a= Verdadeiro Positivo, resultados “liberados” das análises que concordam com os valores de referência.

b= Falso Positivo, resultados “liberados” das análises que discordam dos valores de referência.

c= Falso Negativo, resultados “retidos” das análises que discordam dos valores de referência.

d= Verdadeiro Negativo, resultados “retidos” das análises que concordam com os valores de referência.

Sensibilidade =  $a / (a + c)$

Especificidade =  $d / (b + d)$

Falso Positivo ou Valor Preditivo Positivo =  $a / (a + b)$

Falso Negativo ou Valor Preditivo Negativo =  $d / (c + d)$

Acurácia =  $(a + d) / (a + b + c + d)$

FONTE: O autor (2011)

OpenEpi--Epidemiologic Cal... x

www.openepi.com/OE2.3/Menu/OpenEpiMenu.htm

Expand All | Collapse

OpenEpi

Start Enter Results

## Results

### Diagnostic or Screening Test Evaluation

#### Single Table Analysis

	Positive	Negative	Total
Positive	11031	141	11172
Negative	236	18595	18831
	11267	18736	30003

Parameter	Estimate	Lower - Upper 95% CIs	Method
Sensitivity	97.91%	(97.62, 98.15 <sup>1</sup> )	Wilson Score
Specificity	99.25%	(99.11, 99.36 <sup>1</sup> )	Wilson Score
Positive Predictive Value	98.74%	(98.51, 98.93 <sup>1</sup> )	Wilson Score
Negative Predictive Value	98.75%	(98.58, 98.9 <sup>1</sup> )	Wilson Score
Diagnostic Accuracy	98.74%	(98.61, 98.86 <sup>1</sup> )	Wilson Score
Likelihood ratio of a Positive Test	130.1	(128.3 - 131.9)	
Likelihood ratio of a Negative Test	0.0211	(0.02093 - 0.02128)	
Diagnostic Odds	6164	(4997 - 7604)	
Cohen's kappa (Unweighted)	0.9732	(0.9618 - 0.9845)	
Entropy reduction after a Positive Test	59.41%		
Entropy reduction after a Negative Test	59.45%		
Bias Index	-0.003166		

#### Results from OpenEpi, Version 2, open source calculator--DiagnosticTest

<http://www.openepi.com/OE2.3/DiagnosticTest/DiagnosticTest.htm>  
Source file last modified on 09/21/2010 02:07:02

Print from the browser, or select all or part of the text and then copy and paste to other programs.  
Many browsers have an optional setting to print background colors.

dTree control ©2002-2003  
by [Gerr Lundro](#)

FIGURA 18 - EXEMPLO DO SOFTWARE OPENEPI

FONTE: Livre adaptação de <<http://www.openepi.com/>>. Acesso em 28/01/2011.

As variáveis em estudo foram analisadas quanto à normalidade utilizando os testes de Kolmogorov-Smirnov e Lilliefors. A Tabela 9 mostra os valores de probabilidade para estes testes evidenciando que todos os analitos não apresentam distribuição normal.

A Tabela 9 mostra o resultado da análise de normalidade com todos os parâmetros em estudo.

TABELA 9 - ANÁLISE DE NORMALIDADE PARA AS VARIÁVEIS EM ESTUDO

<b>Grupo “Liberado” (0)</b>	<b>n</b>	<b>max D</b>	<b>K-S</b>	<b>Lilliefors</b>
Sexo	22.719	0,431511	p < 0,01	p < 0,01
Idade	22.719	0,037683	p < 0,01	p < 0,01
Glicemia	22.719	0,053866	p < 0,01	p < 0,01
HDL-colesterol	22.719	0,062827	p < 0,01	p < 0,01
LDL-colesterol	22.719	0,056571	p < 0,01	p < 0,01
Triglicérides	22.719	0,046094	p < 0,01	p < 0,01
Colesterol total	22.719	0,037958	p < 0,01	p < 0,01
<b>Grupo “Retido” (1)</b>	<b>n</b>	<b>max D</b>	<b>K-S</b>	<b>Lilliefors</b>
Sexo	37.287	0,422499	p < 0,01	p < 0,01
Idade	37.287	0,091193	p < 0,01	p < 0,01
Glicemia	37.287	0,259368	p < 0,01	p < 0,01
HDL-colesterol	37.287	0,073192	p < 0,01	p < 0,01
LDL-colesterol	37.287	0,055706	p < 0,01	p < 0,01
Triglicérides	37.287	0,072340	p < 0,01	p < 0,01
Colesterol total	37.287	0,045789	p < 0,01	p < 0,01

K-S, teste de Kolmogorov-Smirnov. p<0,01 foi considerado significativo  
 FONTE: O autor (2011)

Para a comparação entre as variáveis foram utilizados o teste não paramétrico U de Mann-Whitney. Para a análise de correlação foi utilizado o modelo de correlação linear descrito por Pearson e a correlação de Spearman (Spearman *rank R*, não paramétrico), sendo considerado significativo àquelas que apresentaram p<0,001, com a correção de Bonferroni. As frequências entre sexo (masculino=2 e feminino=1) foram comparadas entre os grupos utilizando teste de qui-quadrado com 1 grau de liberdade.

Uma probabilidade (p) inferior a 5% (p<0,05) foi considerada significativa em todas as análises, exceto quando indicado. O programa Statistica para Windows 8.0 foi utilizado em todas as análises.

## 5 RESULTADOS

### 5.1 AMOSTRA

Os resultados dos parâmetros laboratoriais em estudo foram exportados do banco de dados DB2 e convertidos para a planilha Excel. Após a manipulação do arquivo com os recursos da Tabela Dinâmica do Excel e submetido (importado) ao banco de dados MYSQL, os dados foram pareados e utilizados para as análises.

#### 5.1.1 Características do Banco de Dados MYSQL

Uma vez os dados importados no MYSQL, foram trabalhados usando comandos de SQL para obter a classificação (APENDICE B). No APÊNDICE D é mostrado o resultado dos dados no MYSQL.

As características dos dados utilizados neste trabalho são mostradas na Tabela 10. Podem ocorrer mais de um exame (registro) para cada paciente.

**TABELA 10 - CARACTERÍSTICAS GERAIS DO BANCO DE DADOS MYSQL UTILIZADO NO ESTUDO**

<b>Características da amostra</b>	<b>Quantidades</b>
Número de Exames (registros)	60.006
Número de Pacientes	21.553
Mulheres	66%
Homens	34%
Crianças e Adolescentes (<20 anos)	13%
Resultados Liberados (Classe 0)	38%
Resultados Retidos (Classe 1)	62%

Os valores são apresentados em número ou porcentagem

FONTE: O autor (2011)

### 5.1.2 Características da Amostra em Estudo

As características dos registros de pacientes utilizados para análise (n=60.006) contendo informações sobre glicemia em jejum, colesterol total, HDL-colesterol, LDL-colesterol e triglicérides, bem como idade e gênero, e a classificação de liberado ou retido, são mostrados na Tabela 11.

TABELA 11 - CARACTERÍSTICAS DA AMOSTRA EM ESTUDO.

Parâmetros (n=60.006)	Grupos		p
	Liberado n=22.719 (38%)	Retido n=37.287 (62%)	
Idade (anos)	46,9±18,5 48[2-99]	48,7±18,2 52[2-99]	<0,001
Sexo (M/F, %)	32,4/67,6	34,2/65,8	<0,001*
Glicemia (mg/dL)	94,1±11,5 93[50-126]	114,1±53,3 99[10-1.161]	<0,001
HDL-colesterol (mg/dL)	48,1±13,0 47[15-100]	45,4±14,1 44[3-200]	<0,001
LDL-colesterol (mg/dL)	98,5±19,8 101[28-130]	130,8±41,1 134[10-639]	<0,001
Triglicérides (mg/dL)	92,3±29,2 90[15-150]	167,2±77,8 157[6-1.213]	<0,001
Colesterol total (mg/dL)	165,1±26,1 167[64-240]	209,9±48,2 211[17-764]	<0,001

Os valores estão expressos em Média±1-Desvio Padrão e Mediana [Amplitude de Variação]

p, probabilidade teste não paramétrico U de Mann-Whitney

O parâmetro sexo foi calculado o percentual de M/F por grupo (Liberado e Retido)

\*Probabilidade, teste de qui-quadrado

FONTE: O autor (2011)

A Figura 19 mostra histograma da distribuição da idade nos grupos estudados.

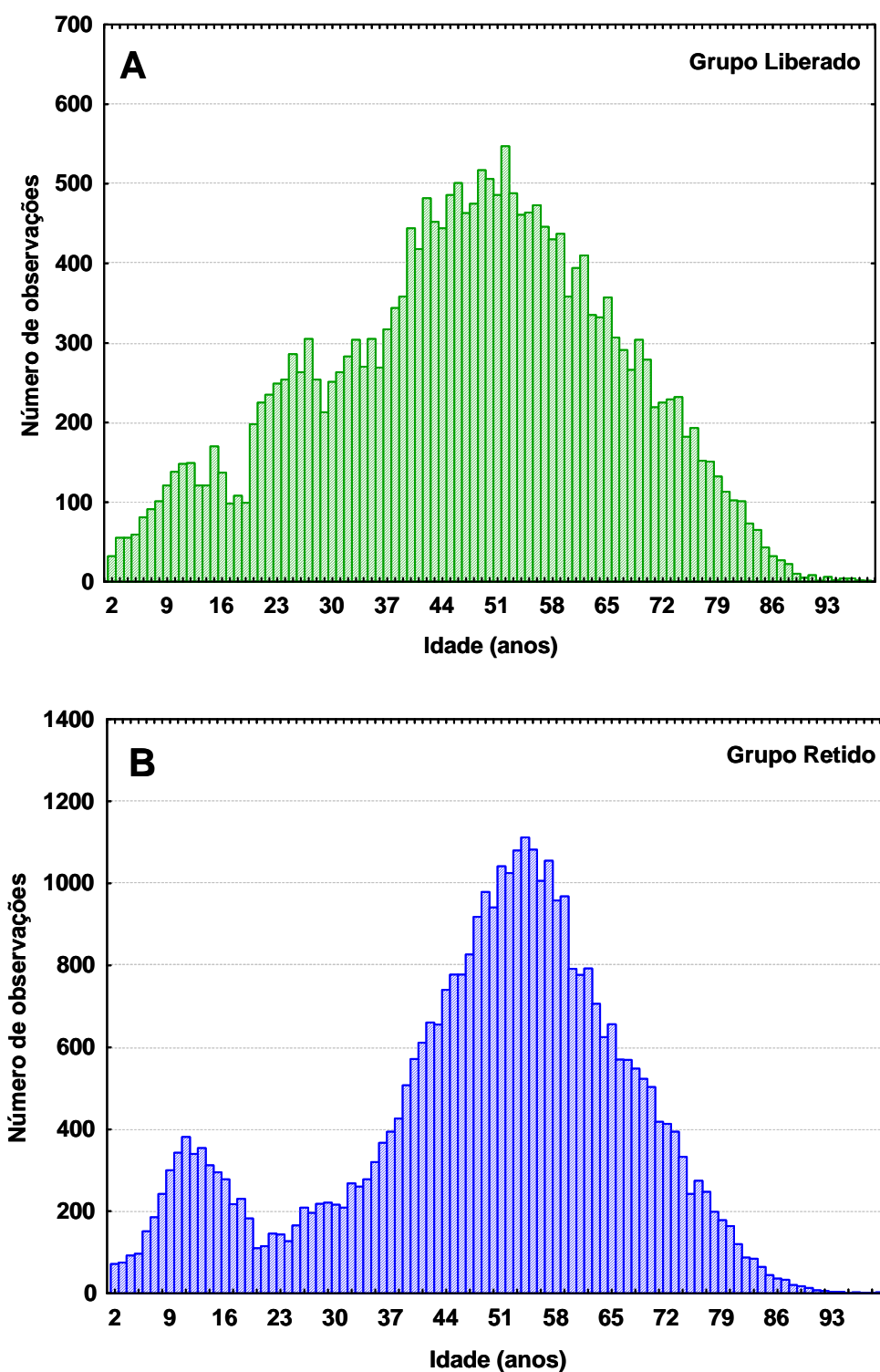


FIGURA 19 - HISTOGRAMA DA VARIÁVEL IDADE NA AMOSTRA CLASSIFICADA NOS GRUPOS EM ESTUDO.

Figura A. Distribuição da idade no grupo liberado (código=0)

Figura B. Distribuição da idade no grupo retido (código=1).

FONTE: O autor (2011)



A Tabela 12 mostra o resultado da análise de correlação linear (Pearson) para os parâmetros Idade, Glicemia em Jejum, HDL-colesterol, LDL-colesterol Triglicérides e Colesterol Total.

**TABELA 12 - ANÁLISE DE CORRELAÇÃO LINEAR (PEARSON) COM OS PARÂMETROS EM ESTUDO.**

<b>N=60.006</b>	<b>Idade</b>	<b>Glicemia</b>	<b>HDL-c</b>	<b>LDL-c</b>	<b>TG</b>
Idade	NA	-	-	-	-
		-	-	-	-
Glicemia	0,1127	NA	-	-	-
	p<0,01		-	-	-
HDL-c	0,0384	-0,0387	NA	-	-
	p<0,01	p<0,01		-	-
LDL-c	0,1307	-0,0206	0,1849	NA	-
	p<0,001	p<0,001	p<0,01		-
TG	0,1418	0,1416	-0,2422	0,1626	NA
	p<0,01	p<0,01	p<0,01	p<0,01	
CT	0,1636	0,0243	0,3794	0,9166	0,3820
	p<0,01	p<0,001	p<0,01	p<0,001	p<0,01

Os valores apresentados são o coeficiente de correlação linear de Pearson (r) e o valor de probabilidade (p). Probabilidade para significância corrigida (Bonferroni) p<0,007. Todos os valores de “r” apresentados são similares aos observados com a correlação não paramétrica de Spearman (Rank R).

HDL-c, HDL-colesterol; LDL-c, LDL-colesterol; TG, triglicérides; CT, colesterol total.

O valor destacado em retângulo pontilhado representa a maior correlação observada.

FONTE: O autor (2011)

## 5.2 ANÁLISE DAS REDES NEURONAIS TREINADAS

Após o treinamento as redes apresentaram acurácia superior a 90%.

A rede MLP apresentou um total de 377 resultados divergentes, obtidos da adição das divergências entre liberados e retidos (141+236), resultando em erro total de 1,2% (377/30.003). Os resultados da matriz de divergência (confusão) são apresentados na Tabela 13.

Os mesmos arquivos de dados foram submetidos ao *software* EasyFAN obtendo acurácia de 92,4% conforme matriz de divergência (confusão) mostrada na Tabela 13. A rede FAN mostrou um total de divergência (confusão) considerando a soma dos resultados divergentes retidos e liberados ( $2.050+232=2.282$ ) de 7,6% ( $2.282/30.003$ ).

TABELA 13 - MATRIZ DE DIVERGÊNCIA PARA AS REDES NEURONAIS MLP E FAN

Classificação por redes	Classificação por valores de referência	
Rede MLP	0-LIBERADO (37,2%)	1-RETIDO (62,8%)
0-LIBERADO	11.031 (36,8%)	141 (0,4%)
1-RETIDO	236 (0,8%)	18.595 (62,0%)
Rede FAN		
0-LIBERADO	10.940 (36,4%)	232 (0,8%)
1-RETIDO	2.050 (6,8%)	16.781 (56,0%)

FONTE: O autor (2011)

Os testes de avaliação, baseados nas matrizes de divergência (confusão), para as redes MLP e FAN são mostrados na Tabela 14. Todas as características para o treinamento e teste foram realizadas com pareamento (C-1 e C-2 em M&M).

TABELA 14 - COMPARAÇÃO DAS REDES NEURONAIS COM ARQUIVOS COMPLETOS

PARÂMETROS	Redes Neurais	
	MLP	FAN
Sensibilidade	97,9%	84,2%
Especificidade	99,3%	98,6%
Valor Preditivo Positivo	98,7%	97,9%
Valor Preditivo Negativo	98,8%	89,1%
Acurácia	98,8%	92,4%

As redes neurais foram treinadas com o arquivo completo (C-1) e testadas com arquivo completo (C-2) como descrito em M&M item 4.1.

Os resultados para os parâmetros de comparação das redes foram realizados utilizando os dados da matriz de divergências (TABELA 13) aplicada ao *software* OPENEPI (FIGURA 18, M&M item 4.6).

FONTE: O autor (2011)

### 5.2.1 Análise da Rede Neuronal FAN com Registros Incompletos

Foram testados para a rede neuronal FAN o arquivo (I-1, M&M) contendo 65.536 registros. Este arquivo foi sequencialmente dividido em dois ( $n=35.000$  e  $n=30.536$  registros), respectivamente para treinamento e teste. Estes arquivos apresentavam os dados sem nenhum tratamento ou filtro, de forma que alguns registros apresentaram reconhecidos valores inconsistentes (*outliers*) como valores negativos, valores registrados como “zero” ou muito elevados apresentado em alguns casos características de erros de digitação. Para a rede FAN não foi possível realizar teste com arquivos com registros superiores a 35 mil linhas. O *hardware* utilizado (notebook Dell Vostro 1310) não apresentou capacidade de processamento para o *software* EasyFAN. O objetivo era estudar e verificar como se comportaria a rede neuronal FAN com quantidades maiores de registros quanto aos parâmetros sensibilidade, especificidade e valores preditivos negativos/positivos. Os resultados de divergência (matriz de confusão) são mostrados na Tabela 15 para o arquivo incompleto testado (I-1, M&M item 4.1).

TABELA 15 - MATRIZ DE DIVERGÊNCIA PARA A REDE FAN COM ARQUIVOS INCOMPLETOS

Rede FAN	Classificação por valores de referência	
n=30.536	0-LIBERADO (50,3%)	1-RETIDO (49,7%)
0-LIBERADO	13.803 (45,2%)	1.564 (5,1%)
1-RETIDO	6.431 (21,1%)	8.738 (28,6%)

Arquivo I-1 (M&M item 4.1) foi dividido em 35.000 para treinamento e 30.536 para teste.  
 FONTE: O autor (2011)

### 5.2.2 Análise da Rede Neuronal MLP com Registros Incompletos

Foram treinados e testados com a rede MLP outros quatro arquivos com maior tamanho amostral (I-1 a I-4, Figura 14, M&M). O arquivo de 65.536 (I-1) foi subdividido em dois arquivos de  $n=35.000$  registros para treinamento e  $n=30.536$  registros para teste. O arquivo de 65.536 (I-1) foi utilizado para treinamento da rede sendo utilizado para teste outro arquivo de 65.536 registros (I-2).

O arquivo de 120.000 (I-3) foi utilizado para treinamento da rede sendo utilizado para teste outro arquivo de 120.000 registros (I-4).

A diferença entre os arquivos de 30.536, 65.536 e 120.000 é que os mesmos possuíam valores negativos, valores zerados para perfil lipídico ou para glicemia em jejum, valores considerado alto, acima da realidade ocasionada por erros de digitação. Os resultados foram os seguintes conforme são mostrados abaixo (TABELA 16) através de matriz de confusão e parâmetros.

**TABELA 16 - MATRIZ DE DIVERGÊNCIA PARA A REDE NEURONAL MLP COM ARQUIVOS INCOMPLETOS**

<b>Rede MLP</b>	<b>Classificação por valores de referência</b>	
<b>n=30.536</b>	<b>0-LIBERADO (50,3%)</b>	<b>1-RETIDO (49,7%)</b>
0-LIBERADO	15.248 (49,9%)	119 (0,4%)
1-RETIDO	270 (0,9%)	14.899 (48,8%)
<b>n=65.536</b>	<b>0-LIBERADO (53,1%)</b>	<b>1-RETIDO (46,9%)</b>
0-LIBERADO	34.405 (52,5%)	382 (0,6%)
1-RETIDO	1.388 (2,1%)	29.361 (44,8%)
<b>N=120.000</b>	<b>0-LIBERADO (56,3%)</b>	<b>1-RETIDO (43,7%)</b>
0-LIBERADO	67.443 (56,2%)	118 (0,1%)
1-RETIDO	314 (0,3%)	52.125 (43,4%)

Arquivos, I-1, I-2, I-3, I-4 (M&M item 4.1)

Arquivo I-1 (M&M item 4.1) foi dividido em 35.000 para treinamento e 30.536 para teste.

FONTE: O autor (2011)

A Tabela 17 mostra o resultado dos parâmetros de desempenho comparativamente entre os arquivos com 30.536 registros, utilizando 500 épocas no treinamento para ambas as redes neurais (MLP e FAN). A mesma quantidade de épocas foi utilizada para as quantidades de 65.536 e 120.000 registros para comparação desses números com a rede neuronal MLP somente.

TABELA 17 - COMPARAÇÃO ENTRE AS REDES MLP E FAN COM ARQUIVOS INCOMPLETOS

PARÂMETROS	Redes Neurais			
	FAN	MLP		
	n=30.536	n=30.536	n=65.536	n=120.000
Sensibilidade	68,2%	98,3%	96,1%	99,5%
Especificidade	84,8%	99,2%	98,7%	99,8%
Valor Preditivo Positivo	89,8%	99,2%	98,9%	99,8%
Valor Preditivo Negativo	57,6%	98,2%	95,5%	99,4%
Acurácia	73,8%	98,7%	97,3%	99,6%

Redes neurais MLP foram treinadas com os arquivos incompletos (I-1 e I-3, M&M) e testadas com arquivos incompletos (I-2 e I-4) como descrito em M&M item 4.1. Ainda foi dividido o arquivo I-1 em dois para treinar e testar as redes neurais MLP e FAN. O valor de 35.000 foi utilizado para treinar e o valor de 30.536 para testar.

FONTE: O autor (2011)

### 5.3 RESULTADOS ANALISADOS POR PROFISSIONAIS ESPECIALISTAS

Os registros divergentes na rede neuronal MLP (n=377), reportados na Tabela 13, foram avaliadas por dois especialistas em Imunoquímica do HC-UFPR e os resultados mostrados na Tabela 18.

TABELA 18 - COMPARAÇÃO ENTRE ANALISTAS HUMANOS COM OS RESULTADOS DIVERGENTES DA REDE MLP

AVALIADORES HUMANOS	CONCORDÂNCIA COM A REDE	COMENTÁRIO DO AVALIADOR
Avaliador A	100%	O avaliador concordou com o resultado da rede, desconsiderando os valores de referência.
Avaliador B	47%	O avaliador considerou os resultados da rede corretos em aproximadamente 50% dos casos. A maior divergência (53%) foi reportada para os valores Retidos (1).

A rede neuronal MLP mostrou divergência em 377 registros (1,2% do total) como mostrada na Tabela 13, sendo que para os registros liberados e retidos, a divergência foi 46,7% (176) e 53,3% (201), respectivamente conforme avaliação do Avaliador B.

FONTE: O autor (2011)

## 6 DISCUSSÃO

As concentrações séricas de glicose e do perfil lipídico são parâmetros laboratoriais que permitem diagnóstico e prognóstico de patologias que afetam grande número de indivíduos na população (ADA, 2010). Como consequência, estes ensaios são responsáveis pela maior demanda tanto na execução quanto na liberação dos resultados do Setor de Imunoquímica de todos os laboratórios clínicos.

A fase analítica, representada pelo processo técnico de quantificação dos analitos teve substancial crescimento e desenvolvimento nas últimas décadas. Analisadores automáticos com crescente capacidade de análise estão disponíveis e tem se tornado progressivamente equipamentos acessíveis a todos os laboratórios.

No entanto, a liberação dos resultados, fase pós-analítica, não apresentou avanços compatíveis com a evolução do processamento das amostras. O interfaceamento entre o analisador e o sistema de liberação de resultados, permite que os profissionais suprimam a emissão de planilhas e listas, bem como transcrições manuais, que favoreciam e facilitavam o aparecimento de erros. Poucos sistemas laboratoriais de emissão de laudos, estão equipados com *softwares* que ajudem, facilitem ou mesmo permitam algum tipo de apoio à decisão, no processo de liberação de resultados.

Os *softwares* de apoio à decisão para laboratórios clínicos disponíveis têm como fundamento a aplicação dos valores de referência do método, ajustados para algumas variáveis como sexo e idade, por exemplo, e informam aos profissionais a recomendação para “liberar”, aqueles dentro da referência e “reter” os que apresentam qualquer valor alterado.

Este trabalho propõe estudar a aplicação de inteligência artificial na forma de redes neurais aplicadas para oferecer aos profissionais do laboratório apoio a decisão na liberação de laudos, também conhecida como segunda opinião. Como as redes neurais “aprendem” são esperadas vantagens neste processo em especial no tratamento de variáveis dependentes.

## 6.1 AMOSTRA

A amostra estudada para arquivos completos (FIGURA 14) é composta dos registros de resultados de exames laboratoriais de 21.553 pacientes da rotina do setor de Imunoquímica do HC-UFPR (TABELA 10). O predomínio de mulheres foi expressivo, na proporção de 2:1, e o grupo de crianças e adolescentes representou 13% do total. Estas características refletem o perfil de atendimento populacional do Hospital de Clínicas da UFPR.

É reconhecido que as mulheres procuram os serviços de atendimento à saúde de forma mais resoluta que os homens (LAURENTI *et al.*, 2005). Este achado é comprovado no presente estudo.

As características laboratoriais da amostra estudada (TABELA 11) confirmam que o processo de classificação realizado por valores de referência foi eficaz. Todo o grupo de exames classificado como “liberado” encontra-se dentro dos limites máximos estabelecidos pelas regras utilizadas (TABELA 5). Entre os analitos estudados, triglicérides e glicemia em jejum apresentaram as maiores variações. A determinação de triglicérides é conhecida por propiciar em alguns pacientes valores muito elevados (PICHETH *et al.*, 2001), o que justifica o resultado observado no presente estudo que mostraram variações de 6 mg/dL a 1.213 mg/dL (TABELA 11).

A determinação de glicemia mostrou uma variação no grupo “retido” de 10 mg/dL a 1.161mg/dL. Esta ampla variação pode ser explicada pelo atendimento de pacientes graves, com diabetes descontrolado ou em crise de hipoglicemia, processos patológicos rotineiros em um hospital do porte e da complexidade de atendimento como o HC-UFPR.

Todos os valores médios do grupo “retido” foram significativamente diferentes do grupo “liberado” ( $p < 0,05$ ). Este achado era esperado e decorrente da classificação da amostra. A distribuição entre homens e mulheres nos grupos estudados foi diferente, com maior proporção de mulheres no grupo “liberado”.

A porcentagem de resultados retidos (62,8%) foi expressivamente superior a de liberados (37,2%) aplicando os critérios de referência do presente estudo (TABELA 5). Pacientes com resultados alterados na proporção de 2:1 em relação aqueles com resultados “normais” era esperado, uma vez que o atendimento hospitalar é reservado principalmente àqueles que não tiveram seus problemas de saúde resolvidos em postos de saúde e situações de emergência.

Este trabalho foi o primeiro a estabelecer esta proporção com base científica para os parâmetros estudados no HC-UFPR, o terceiro maior hospital universitário do Brasil. Volanski (2011) estudando parâmetros de Urinálise também observou uma proporção de 2:1 entre amostras “alteradas” e “normais” para uma população ambulatorial de Curitiba. Não foram identificados outros estudos para comparação com os dados produzidos nesta dissertação, o que reforça a necessidade de ampliar o conhecimento do padrão laboratorial das instituições hospitalares no Brasil.

A elevada proporção de exames alterados propicia que os profissionais que atuam na liberação de resultados tenham uma sobrecarga pela necessidade de análise detalhada de grande volume de dados. Esta observação reforça a necessidade de investimentos e novos sistemas de apoio à decisão, que minimizem o trabalho destes profissionais.

Uma análise da distribuição da idade entre os grupos “liberado” e “retido” foi realizada para ampliar o conhecimento da amostra estudada (FIGURA 19). O grupo “liberado” apresenta uma distribuição da idade próxima à “normal”, com pico de maior frequência em 55 anos (FIGURA 19A). Em contrapartida, o grupo “retido” mostra uma distribuição da idade com característica bimodal. Neste grupo, um pico entre jovens com idade de 12 anos e em adultos com 55 anos foi observado (FIGURA 19B). Em ambos os grupos estudados, uma faixa de 44 a 58 anos representa o maior número de atendimentos. Este perfil representa uma característica própria, específica do atendimento do HC-UFPR.

Os parâmetros estudados são altamente correlacionados (TABELA 12). Este achado está em concordância com outros estudos (MA *et al.*, 2006; LADEIA *et al.*, 2006). A elevada correlação era esperada, uma vez que os elementos do perfil lipídico estudados, bem como alterações no metabolismo de carboidratos são relacionados. A maior correlação observada foi entre as determinações de colesterol total e LDL-colesterol ( $r=0,9166$ ;  $p<0,001$ ). Este achado era esperado, uma vez que os valores de LDL-colesterol foram estimados pela equação de Friedewald, que contempla os valores de colesterol total como elemento central no cálculo (M&M 4.3). Os resultados apresentados na Tabela 12 são aqueles obtidos pela correlação linear de Pearson. Como o tamanho amostral pode afetar o resultado da correlação, também foi realizado o ensaio não paramétrico de Spearman (*Rank R*). Este apresentou valores de correlação similares aos apresentados no estudo, confirmando a elevada correlação.



A análise de correlação foi realizada para comprovar o grau de correlação, fator que é passível de afetar o desempenho de redes neuronais, estudadas neste projeto. Na MLP as características são fortemente dependentes e na FAN são independentes.

## 6.2 ANÁLISE DAS REDES NEURONAIS

As redes neuronais FAN e MLP, escolhidas para o estudo, foram estudadas em duas condições distintas. A primeira, analisando arquivos completos, caracterizados pela presença de todos os elementos (analitos e dados demográficos) em todos os registros. E a segunda, estudando arquivos incompletos, aonde todos os registros continham dados demográficos (idade e sexo) e valores para os analitos variados de um a cinco, mimetizando a liberação real do serviço do HC-UFRP.

### 6.2.1 Análise das Redes Neuronais com Arquivos Completos

Uma base de dados de 60.006 registros completos foi subdividida em arquivos iguais ( $n=30.003$ ) de treino e teste (FIGURA 14, C-1 e C-2). Neste estudo, a rede FAN apresentou divergência (TABELA 13) cerca de 6 vezes superior à rede MLP (7,6% e 1,2%, respectivamente). A rede FAN teve como principal fonte de divergência (6,8%) resultados “retidos” quando comparado com a liberação por valores de referência. Comparativamente o desempenho da rede MLP foi superior, com divergência de 0,8% (TABELA 13) para os valores “retidos”.

Quando as características das duas redes neuronais estudadas foram comparadas (TABELA 14) todos os parâmetros de análise para a rede MLP foram superiores aos apresentados pela rede FAN. Entre estes parâmetros, a sensibilidade, valor preditivo negativo e acurácia apresentaram as maiores diferenças. A sensibilidade reflete a característica da rede neuronal estudada em apresentar resultados compatíveis (“liberados”) com o sistema de liberação por valores de referência, utilizados como padrão preliminar de análise, neste estudo. A maior divergência entre as redes neuronais estudadas foi quanto à sensibilidade. A rede MLP (97,8%) comparada à rede FAN (84,2%) apresentou diferença de 13,6 pontos percentuais. As análises destes parâmetros substanciam que a rede neuronal MLP comparada à rede FAN, com amostras completas apresenta melhor desempenho.

### 6.2.2 Análise das Redes Neurais com Arquivos Incompletos

O desempenho da rede neuronal FAN estudado com arquivos incompletos (M&M, Figura 14, I-1) foi expressivamente inferior ao observado com arquivos completos (TABELAS 13 e 14). Neste estudo, a rede FAN apresentou divergência de 26,2% (7.995/30.536, Tabela 15) significativamente superior ao observado para arquivos completos (7,6%). Portanto, é possível concluir que a rede neuronal FAN não processa de forma eficiente arquivos incompletos que mimetizam a realidade da rotina laboratorial. A eventual aplicação desta rede em banco de dados como o disponível no HC-UFPR necessitaria um tratamento prévio dos dados para subsequente utilização da rede.

Para a rede FAN não foi possível no presente estudo, avaliar seu desempenho em arquivos com base de dados superiores a cerca de 65 mil registros. Este efeito foi atribuído às características do *software* e do *hardware* utilizados. Com relação ao *software EasyFAN*, o mesmo foi desenvolvido em 2006 sendo que nesse ano ainda trabalhava baseado em arquivos textos e planilhas do Microsoft Excel 2003 que tinha uma limitação de 65.536 linhas ou registros. O *hardware* utilizado na dissertação não conseguiu processar quantidades acima de 35 mil registros sugerindo dessa forma um hardware com configuração superior.

Quando a rede neuronal MLP foi submetida a amostras incompletas apresentou excelente desempenho, com divergências de 1,2% (389/30.536); 2,7% (1.770/65.536) e 0,36% (432/120.000), respectivamente para arquivos de 30.536, 65.536 e 120.000 registros (TABELA 16).

Comparados os desempenhos das redes neurais com os parâmetros estudados (TABELA 17), é notória a superioridade da rede neuronal MLP. Enquanto esta apresentou valores superiores a 95,5% para todos os parâmetros, a rede FAN mostrou o valor mais elevado para valor preditivo positivo de 89,8% e todos os demais parâmetros abaixo de 85%.

A sensibilidade para a rede FAN com arquivos incompletos inferior a 70% comparada com valores superiores a 96% para a rede MLP reforça que a rede FAN não é recomendada para esta amostragem, diferentemente da rede neuronal MLP.

O desempenho da rede neuronal MLP foi similar com o uso de arquivos completos e incompletos (TABELAS 14 e 17). Outro elemento relevante a ser ressaltado é que o número de registros de treinamento (tamanho da amostra) não

afetou o desempenho da rede, como observado para arquivos incompletos de 30.536, 65.536 e 120.000 registros (TABELA 17). Estes resultados reforçam a efetividade da rede neuronal MLP em processar adequadamente dados completos e incompletos, mostrando a flexibilidade e versatilidade deste processo, em comparação com a rede neuronal FAN.

Portanto, a rede neuronal MLP tem potencial para aplicação a registros similares aos disponibilizados no banco de dados do HC-UFPR, sem necessidade de tratamento prévio dos dados e com elevada eficiência de classificação após treinamento. A rede neuronal MLP é recomendada para futuras aplicações em estudos similares ao apresentado neste trabalho.

A aplicação de redes neurais para apoio a decisão no laboratório clínico é um campo recente. Exceto pelo trabalho de Volanski (2011), não foram detectados na literatura pesquisada trabalhos científicos similares para comparação com os resultados obtidos no presente trabalho. As redes neurais vêm sendo extensivamente estudadas e aplicadas na área da saúde, para o setor de diagnóstico de imagem, também como o objetivo de fornecer um apoio à decisão clínica. Neste contexto, comparando os resultados apresentados nesta dissertação para a rede neuronal MLP, que apresentou o melhor desempenho, com acurácia superior a 97% em todas as análises realizadas, os resultados são comparáveis aos descritos para análise do eletrocardiograma com rede neuronal MLP com acurácia de 88,7%; 88,5% e 91,0%, respectivamente descritos por Soares e Nadal (1999), Dourado, Souza e Oliveira (2008) e Rosario, Monteiro e Gomes (2008).

Um arquivo contendo 377 registros de divergência entre a rede MLP comparado com os critérios de valores de referências, foi submetido a dois avaliadores humanos, especialistas na área (TABELA 18). Um dos avaliadores concordou com os resultados emitidos pela rede em 100% e outro em apenas cerca de 50%. Esta divergência entre avaliadores humanos reforça a dificuldade da tomada de decisão e a variação expressiva resultante do processo. Os valores “retidos” foram para o avaliador B o principal elemento de divergência. A análise realizada por especialistas humanos engloba no processo de liberação e avaliação dos resultados características como a experiência profissional, critério relevante, mas com elevado grau de subjetividade, o que explica as divergências reportadas entre os analistas. Uma análise com maior número de analistas humanos é recomendada para futuros estudos envolvendo redes neurais.

### 6.3 PERSPECTIVAS FUTURAS

O potencial para aplicação de redes neuronais no laboratório clínico é crescente com o desenvolvimento desta tecnologia. Também as novas necessidades do laboratório em especial no processo de automação para a liberação de resultados necessitam de novas ferramentas de informática e a inteligência artificial tem destaque neste contexto.

É possível supor que redes neuronais possam oferecer um caminho seguro para os *softwares* de apoio a decisão. A partir da análise de dados altamente relacionados, será inevitável no futuro a integração entre todos os setores do laboratório bem como a comunicação com dados de outras redes neuronais envolvidas nos diagnósticos clínicos e de imagem. O uso amigável e a confiabilidade destes *softwares* são elementos essenciais para a aceitabilidade destas novas tecnologias pelos profissionais da saúde.

A complexidade do processo descrito oferece uma oportunidade e estímulo para o desenvolvimento de novos *softwares* com uso redes neuronais, com design específico para área da saúde, bem como a criação de processos inovadores para compatibilizar esta integração de indicadores de saúde/doença para benefício do ser humano.

Como novas perspectivas de trabalho para ampliar o desenvolvimento do projeto, avaliar o desempenho da rede neuronal FAN com mais épocas de treinamento, é relevante. A sugestão é treinar a rede FAN com mais de 500 épocas e pelo menos por 12 horas. A rede neuronal FAN tem um comportamento diferente da MLP. A rede neuronal MLP trabalha com características dependentes e a FAN com características independentes. Também novas avaliações do tamanho amostral no treinamento e teste das redes seriam pertinentes.

Também é relevante no futuro avaliar o desempenho das redes neuronais em análises de conjuntos de dados que permitam a classificação e estratificação de risco associado à doença cardiovascular e às complicações do diabetes. Este objetivo será contemplado integrando os resultados laboratoriais aos dados clínicos dos pacientes, e a aplicação de inteligência artificial como elemento de decisão para os critérios que forem estabelecidos.

## 7 CONCLUSÕES

- O banco de dados desenvolvido em MYSQL permitiu a classificação e manipulação dos registros com eficiência (produtividade).
- Quando as variáveis: sexo, idade, glicemia em jejum e perfil lipídico (colesterol total, HDL-colesterol, LDL-colesterol e triglicérides) foram classificadas utilizando valores de referência, observou-se que aproximadamente 62% da amostra estudada apresentou valor fora da referência em pelo menos uma das variáveis.
- As redes neurais MLP e FAN, quando comparadas com arquivos completos, apresentaram desempenho satisfatório (acurácia superior a 90%).
- A rede neuronal FAN perde desempenho quando treinada e testada com arquivos incompletos para a amostra estudada.
- A rede neuronal MLP apresentou desempenho superior à rede FAN quando analisada com arquivos completos ou incompletos.
- O tamanho amostral utilizado para treinamento e teste não afeta o desempenho da rede neuronal MLP.
- Resultados divergentes da rede neuronal MLP avaliados por especialistas humanos evidenciaram que os ensaios com valor alterado foram o principal elemento de inconsistência.
- A rede neuronal MLP é recomendada para outros estudos com desenho amostral semelhante e apresenta potencial para aplicação no laboratório clínico como suporte a decisão na liberação de resultados.

## REFERÊNCIAS

ABBOTT. Laboratories. **Architect ci8200 interactive demonstration**. Disponível em: <<http://www.abbottdiagnostics.co.uk/>>. Acesso em: 30 set. 2010

ADA. Diagnosis and classification of diabetes *mellitus*. **Diabetes care**. v. 33 Supl 1, p. S62-9. 2010.

AGOSTI, C.; BAYS, C.; MORENO, M. **Application of artificial neural networks in the mammographic diagnosis**. 13 Congresso Mundial de Mastologia, Recife/Pernambuco/Brasil, p. 174-174, 2004.

AMBRÓSIO, P. E. . **Redes neurais artificiais no apoio ao diagnóstico diferencial de lesões intersticiais pulmonares**. Ribeirão Preto: FFCLRP/USP, 2002.

AUER, J.; WEBER, T.; BERENT, R.; LASSNIG, E.; MAURER, E.; LAMM, G.; KVAS, E.; EBER, B. Obesity, body fat and coronary atherosclerosis. **Int. J. Cardiol.**, v.98, p. 227-235, 2005.

BISHOP, C. M.; **Neural networks for pattern recognition**; Oxford University Press; 1995.

BRAGA, A. P.; CARVALHO, A. C. P. L. F.; LUDEMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: LTC, 2000.

CHAMBERLIN, D. D., ASTRAHAN, M. M., BLASGEN, M. W., GRAY, J. N., KING, W. F., LINDSAY, B. G., LORIE, R., MEHL, J. W., PRICE, T. G., PUTZOLU, F., SELINGER, P. G., SCHKOLNICK, M., SLUTZ, D. R., TRAIGER, I. L., WADE, B. W., and YOST, R. A. **A history and evaluation of System R**. **Commun. ACM** 24, 10 (Oct. 1981), 632-646

COX, E. **The Fuzzy Systems Handbook: A Practitioner's Guide to Building, Using, and Maintaining Fuzzy Systems**. Academy Press Inc., 1994.

DEAN, A.G., SULLIVAN, K.M.; SOE, M.M. **OpenEpi: Open Source Epidemiologic Statistics for Public Health, Version 2.3.1**. Updated 2010/19/09. Disponível em: <<http://www.openepi.com>>. Acesso em: 18 jan. 2011.

DOURADO, H.; SOUZA, J.; OLIVEIRA, O. **Análise de ECG para detecção de infarto anterior com a utilização de wavelet**. 2008. TCC (Bacharelado em Ciência da Computação) - Faculdade Ruy Barbosa, Salvador, 2008.

FALQUETO, J.; SOUZA, C. G.; TOSTES, L. M. **O uso de redes neurais artificiais no diagnóstico de doenças reumatológicas**. Florianópolis: UFSC, 2003.

FERNEDA, E. **Redes neurais e sua aplicação em sistemas de recuperação de informação**. *Ci. Inf.* [online]. 2006, vol.35, n.1, pp. 25-30. ISSN 0100-1965.

HAWKER, C. D. Laboratory automation: total and subtotal. **Clin. Lab. Med.**, v. 27, n. 4, p. 749–770, 2007.

HAYKIN, S. **Redes neurais: princípios e prática**. 2. Ed. Porto Alegre: Bookman, 2001.

HC, Site Oficial do. **Relatório de Atividades – 2009**. Disponível em: <<http://www.hc.ufpr.br>>. Acesso em: 28 jun. 2010.

IBM, Site Oficial da. **Download DB2 Express-C**. Disponível em: <<http://www-01.ibm.com/software/data/db2/express/download.html>>. Acesso em: 18 jul. 2010.

KUSTER C V, IGNACIO F A, LENFERS F P, GARRETT L F V, ZOTTO S. **EasyFan**. 2006. Trabalho de Conclusão de Curso. (Graduação em Tecnólogo em Informática) - Universidade Federal do Paraná.

LADEIA, A. M.; ADAN, L.; COUTO-SILVA, A. C.; HILTNER, A.; GUIMARÃES, A. C. Lipid profile correlates with glycemic control in young patients with type 1 Diabetes mellitus. **Prev. Cardiol.** v.9 p.82–88, 2006.

LAURENTI, R; JORGE, M. H.P. M. E GOTLIEB, S. L. D. Perfil epidemiológico da morbi-mortalidade masculina. Epidemiological profile off men: morbidity and mortality. **Ciência e Saúde**. v.10, n.1, p.35-46, 2005.

LEVENBERG, K. **A method for the solution of certain non-linear problems in least squares**. The Quarterly of Applied Mathematics 2: (1944) 164–168.

MA, Y.; LI Y., CHIRIBOGA, D. E., OLENDZKI, B. C., HEBERT J. R., KATHERINE, W. L. LEUNG, HAFNER, A. R., OCKENE, I. S. **Association between carbohydrate intake and serum lipids**. **J. Am. Coll. Nutr.**, v. 25, n. 2, p.155–163, 2006;

MASLAKOWSKI, M. **Aprenda em 21 dias mysql**. Rio de Janeiro: Editora Campus, 2000.

MATHWORKS, Site Oficial do. **Accelerating the pace of engineering and science**. Disponível em: <<http://www.mathworks.com>>. Acesso em: 28 dez. 2010.

MATSUMOTO, T.; UEDA, Y.; KAWAJI, S. **A software system for giving clues of medical diagnosis to clinician**. In: Proc. 15th IEEE Symp. Computer-Based Medical Systems (CBMS), p. 65–70, 2002.

MCCULLOCH, W. S.; PITTS, W. H. **A logical calculus of the ideas immanent in nervous activity**. **Bulletin of Mathematical Biophysics**, n. 5, p.115-133, 1943.

MINSKY, M. L.; PAPPERT, S. **Perceptron: an introduction to computational geometry**. Cambridge: MIT Press, 1969.

MONAT, A. S.; MARTINS, L. W. C. S.; ASSIS, J. **Aplicações de redes neurais para o diagnóstico diferencial da doença meningocócica**. In: IX Congresso Brasileiro de Informática em Saúde, 2004, Ribeirão Preto, SP. Site do IX CBIS, 2004.

MOTTA, V.T. **Bioquímica clínica para laboratório: princípios e interpretações**. 5. Ed. Caxias do Sul: Medbook, 2009.

NAUCK, D., KRUSE, R. **What are Neuro-Fuzzy classifiers?** In Proc. of Seventh International Fuzzy Systems Association World Congress IFSA'97, Vol. 4, pages 228-233, Academia Prague, 1997.

NCEP ATP III. National Cholesterol Education Program - **Adult Treatment Panel III** - NIH Publication No. 02-5215 Sep 2002.

NHLBI, Site Oficial do. **U.S. Department of Health & Human Services**. Disponível em: < [http://www.nhlbi.nih.gov/guidelines/obesity/prctgd\\_c.pdf](http://www.nhlbi.nih.gov/guidelines/obesity/prctgd_c.pdf)>. Acesso em: 28 nov. 2010.

OMS, Organização Mundial da Saúde, **Prevenção de doenças crônicas. Um investimento vital - 2005**. Acesso em 12 out. 2010. Disponível em <[http://www.who.int/chp/chronic\\_disease\\_report/part1\\_port.pdf](http://www.who.int/chp/chronic_disease_report/part1_port.pdf)>.

PAL, S.K., A. SKOWRON (Eds.), **Rough-Fuzzy Hybridization: A New Trend in Decision Making**, Springer, Heidelberg, 1999.

PAL, S.K., S. MITRA, **Neuro-Fuzzy Pattern Recognition: Methods in Soft Computing**, Wiley, New York, 1999.

PESCE, A. J.; KAPLAN, L. A. **Methods in clinical chemistry**. St. Louis: Mosby. 1987. xxii, 1366 p.

PICHETH; G., SCARTEZINI, M. ; ALCÂNTARA, V. M ; MONTEIRO, R. A. ; MARIANO, C. ; JAWORSKI, M. C. G. ; KIKUTI, M. ; PINTO, A. P. ; GRANATO, E. S. **Quanto podem variar duas determinações sucessivas de triglicérides?** LAES/HAES, São Paulo, v. 6, n. 128, p. 174-190, 2001

PHILLIPS, G.B.; PINKERNELL, B.H.; JING, T-Y. **Are major risk factors for myocardial infarction the major predictors of degree of coronary artery disease in men?** *Metabolism*, v.53, n.3, p.324-329, 2004.

QUINLAN, J. R. Induction of decision trees. **Machine learning**, v. 1, n. 1, p. 81–106, 1986.

RAITZ, R. T. **Free Associative Neurons - FAN Uma Abordagem para Reconhecimento de Padrões**. Dissertação de Mestrado. UFSC. 1997.

RAITZ, R. T. **FAN 2002: Um Modelo Neuro-Fuzzy para Reconhecimento de Padrões**. Tese de Doutorado. UFSC. 2002.

ROSÁRIO, D.; MONTEIRO, D.; GOMES, M. **Software inteligente para detecção de infarto anterior**. 2008. TCC (Bacharelado em Ciência da Computação) - Faculdade Ruy Barbosa, Salvador, 2008.



ROSEGREN, A.; WELIN, L.; TSIPOGIANNI, A.; WILHELMSEN, L. Impact of cardiovascular risk factors on coronary heart disease and mortality among middle age diabetic men. A general population study. **BMJ**, v.299, p.1127-1131, 1989.

ROSENBLATT, F. **The Perceptron: a probabilistic model for information storage and organization in the brain**. Psychological Review, v. 65, p. 386-408, 1958.

RUSSELL, S. J.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. Prentice Hall, 2009.

SBC, III Diretrizes Brasileiras Sobre Dislipidemias e Diretriz de Prevenção da Aterosclerose. **Arq. Bras. Cardiol.** V. 77; supl. III, 2001. p.5-7;13-19;26-27;32-35;42.

SBC, IV Diretrizes Brasileiras Sobre Dislipidemias e Prevenção da Aterosclerose. Departamento de Aterosclerose da Sociedade Brasileira de Cardiologia. **Arq. Bras. Cardiol.** V. 88 (supp I), 2007.

SCARTEZINI, M.; PICHETH, G.; SALGADO FILHO, W.; IHARA, S. S. M.; PINTO, L. E. S. A.; TORRES, K. P.; MARTINEZ, T. L.R. Capítulo I, Metabolismo dos lípidos e lipoproteínas. In: MARTINEZ, T. R. L. **Manual de condutas clínicas em Dislipidemias**. Rio de Janeiro: MedLine, 2003: 392p. p. 21-33. ISBN 85-89896-01-3.

SILVA E., OLIVEIRA, A. **Dicas para Configuração de Redes Neurais**. Disponível em: <[http://equipe.nce.ufrj.br/thome/grad/nn/mat\\_didatico/dicas\\_configuracao\\_rna.pdf](http://equipe.nce.ufrj.br/thome/grad/nn/mat_didatico/dicas_configuracao_rna.pdf)> Acesso em: 4 fev. 2011.

SIVATHASAN, S.; CECELJA F.; BALACHANDRAN W. **ECG Diagnosis using Neural Network and Fuzzy Expert System**. In: IEEE INSTRUMENTATION AND MEASUREMENT TECHNOLOGY CONFERENCE, 17, 2000, Brunel University, U.K. Proceedings. 2000, U.K: Brunel University, 2000. p. 988-992.

SOARES, P. P. S.; NADAL, J. **Aplicação de uma Rede Neural Feedforward com Algoritmo de Levenberg-Marquardt para Classificação de Alterações do Segmento ST do Eletrocardiograma**. In: BRAZILIAN CONFERENCE ON NEURAL NETWORKS, 4, 1999, São José dos Campos, p. 384-389.

STAMLER, J.; VACCARO, O.; NEATON, J.D.; WENTWORTH, D. Diabetes and other risk factors – 12 year cardiovascular mortality from the Multiple Risk Factor Intervention Trial. **Diabetes Care**, v.16, p.434-444, 1993.

STATSOFT; Site Oficial do. **The Creators of Statistica**. Disponível em: <<http://www.statsoft.com>>. Acesso em: 28 jan. 2011.

TAKAISHI, H.; TANIGUCHI, T.; FUJIOKA, Y.; ISHIKAWA, Y.; YOKOYAMA, M. Impact of increasing diabetes on coronary artery disease in the past decade. **J. Atheroscler. Thromb.**, v.11, p.271-277, 2004.

VOLANSKI, W. **Desenvolvimento e Validação de Sistema de Apoio à Decisão em Urinálise com Inteligência Artificial Utilizando Redes Neurais**. Dissertação de Mestrado. UFPR, 2011.

WAMPSEVER, Site Oficial do. **Apache, MySQL, PHP on Windows**. Disponível em: <<http://www.wampserver.com/en/>>. Acesso em: 18 dez. 2010.

ZADEH, L.A. **A New Direction in AI: Toward a Computational Theory of Perceptions**, AI Magazine 22 (2001) 73–84.

ZADEH, L.A. **Fuzzy Logic, Neural Networks, and Soft Computing**, Comm. ACM 37 (1994) 77–84.

## APÊNDICES

APÊNDICE A – INSTRUMENTO DE COLETA DE DADOS .....	90
APÊNDICE B – COMANDOS SQL USADOS NO PROJETO .....	92
APÊNDICE C – SCRIPTS EM MATLAB USADOS NO PROJETO .....	94
APÊNDICE D – EXEMPLO DO BANCO DE DADOS NO MYSQL .....	96
APÊNDICE E – EXEMPLO DA INTERFACE GRÁFICA: REDE NEURONAL MLP....	97
APÊNDICE F – EXEMPLO DA INTERFACE GRÁFICA: REDE NEURONAL FAN....	98

## APÊNDICE A – INSTRUMENTO DE COLETA DE DADOS

O objeto utilizado para pesquisa da dissertação de mestrado, exames laboratoriais clínicos, passou por vários processos. A seguinte rotina foi necessária para se alcançar o resultado desejado. Uma observação importante: Não foram utilizadas informações do paciente, tais como: REGISTRO, NOME, ENDEREÇO, CPF e afins; mantendo todo o sigilo do paciente. Resumidamente, os seguintes passos foram seguidos:

1. O médico requisita para o paciente internado ou atendido em ambulatório, exames através do sistema informatizado no HC chamado SIH na opção do módulo de Laboratório.
2. Após a requisição, é coletado amostra do paciente e em seguida os exames dão entrada na seção. Nesse momento é registrado esse controle no sistema SIH tornando o exame no estado “EM ANDAMENTO” no sistema SIH. No caso do projeto será a seção de Imunoquímica pertencente à Unidade de Apoio a Diagnóstico responsável pela realização dos mesmos.
3. As amostras são submetidas ao equipamento chamado Architect da Abbott® que fará a análise química. O *software* instalado nesse equipamento gera resultados através de relatório que posteriormente é utilizado para digitar os valores do resultado do exame no sistema SIH.
4. Uma vez registrado os resultados no SIH, será realizada pesquisa no banco de dados chamado DB2 através de comandos SQL.
5. Os dados utilizados para pesquisa no projeto da dissertação serão os seguintes:
  - a. Sexo do paciente;
  - b. Idade do paciente;
  - c. Valor dos seguintes exames:
    - Glicemia em Jejum;
    - HDL-c;
    - LDL-c;
    - Triglicérides;
    - Colesterol Total;
  - d. Estes dados serão as características utilizadas para a rede neuronal, portanto serão 7 características.

6. Esses dados selecionados no banco de dados DB2 serão convertidos em arquivo no formato Microsoft Excel e trabalhados de forma a serem alinhados lado a lado (pareados) para cada paciente.
7. Após foi modelado um banco de dados no MySQL e posteriormente importado os dados.
8. Nesse ponto ocorre à classificação dos exames através de comandos SQL, definida com base nos valores de referência pela literatura.
9. Após a classificação, os dados foram exportados para formato texto.
10. Este arquivo texto foi dividido em 2 partes: um para treinamento, outro para teste.
11. Os dois arquivos no formato texto foram submetidos à rede neuronal MLP (scripts no MATLAB) e FAN (EasyFAN).
12. Concluiu-se após esta etapa que a rede neuronal MLP teve um desempenho melhor por conta das características serem dependentes.
13. Obteve-se Matriz de confusão e dados que a rede MLP classificou como divergentes foram extraídos para análise por profissional especialista.
14. Após análise dos dados do profissional especialista espera-se constatar:
  - a. Necessita de profissional específico para liberação ou não;
  - b. Linha de corte (*cutoff*) não ser suficiente para definir a liberação dos exames;
  - c. A característica que foi preponderante para a rede neuronal “confundir” a classificação (o valor de LDL-c foi decisivo).
15. Com base nas respostas do item 14 realizar discussões, sugestões e conclusões.

## APÊNDICE B – COMANDOS SQL USADOS NO PROJETO

### 1) Comandos para extração dos resultados de exames do DB2 no HC:

#### **262-GLICOSE EM JEJUM - Método: 135 - HEXOQUINASE/G6PDH**

```
select t2.registro, t2.sexo, year(date(t1.dthrllibrou)-t2.datanasc) as idade,
date(t1.dthrllibrou) as liberacao, t1.cdexame, t1.resnum
from lbresdeterm t1, ihcadastro t2, lbrequisicao t4
where t1.requisicao=t4.requisicao and t1.registro=t2.registro and t1.registro=t4.registro and
t2.registro=t4.registro and t1.requisicao>0 and t1.cdexame=262 and t1.cdamostra=2 and
t1.cddeterm=235 and t1.registro>0 and t4.requisicao>0 and t4.cdconvenio is null
```

#### **270-COLESTEROL HDL - Método: 134 - ENZIMATICO/COLORIMETRICO**

```
select t2.registro, t2.sexo, year(date(t1.dthrllibrou)-t2.datanasc) as idade,
date(t1.dthrllibrou) as liberacao, t1.cdexame, t1.resnum
from lbresdeterm t1, ihcadastro t2, lbrequisicao t4
where t1.requisicao=t4.requisicao and t1.registro=t2.registro and t1.registro=t4.registro and
t1.requisicao>0 and t1.cdexame=270 and cdamostra=2 and t1.cddeterm=175 and t4.cdconvenio is
null
```

#### **271-COLESTEROL LDL - Método: 635 - CALCULO(FORMULA DE FRIEDEWALD)**

```
select t2.registro, t2.sexo, year(date(t1.dthrllibrou)-t2.datanasc) as idade,
date(t1.dthrllibrou) as liberacao, t1.cdexame, t1.resnum
from lbresdeterm t1, ihcadastro t2, lbrequisicao t4
where t1.requisicao=t4.requisicao and t1.registro=t2.registro and t1.registro=t4.registro and
t1.requisicao>0 and t1.cdexame=271 and cdamostra=2 and t1.cddeterm=176 and t4.cdconvenio is
null
```

#### **280-TRIGLICÉRIDES - Método: 633 - GPO (TRIGLICERIDEOS)**

```
select t2.registro, t2.sexo, year(date(t1.dthrllibrou)-t2.datanasc) as idade,
date(t1.dthrllibrou) as liberacao, t1.cdexame, t1.resnum
from lbresdeterm t1, ihcadastro t2, lbrequisicao t4
where t1.requisicao=t4.requisicao and t1.registro=t2.registro and t1.registro=t4.registro and
t1.requisicao>0 and t1.cdexame=280 and cdamostra=2 and t1.cddeterm=202 and t4.cdconvenio is
null
```

#### **285-COLESTEROL TOTAL - Método: 134 - ENZIMATICO/COLORIMETRICO**

```
select t2.registro, t2.sexo, year(date(t1.dthrllibrou)-t2.datanasc) as idade,
date(t1.dthrllibrou) as liberacao, t1.cdexame, t1.resnum
from lbresdeterm t1, ihcadastro t2, lbrequisicao t4
where t1.requisicao=t4.requisicao and t1.registro=t2.registro and t1.registro=t4.registro and
t1.requisicao>0 and t1.cdexame=285 and cdamostra=2 and t1.cddeterm=174 and t4.cdconvenio is
null
```

## 2) Criação da tabela chamada “exames” no banco de dados MySQL:

```
CREATE TABLE `exames` (
  `REGISTRO` int(9) NOT NULL,
  `SEXO` char(1) NOT NULL,
  `IDADE` int(3) NOT NULL,
  `RESULTADO` date NOT NULL,
  `GLICEMIA` int(4) NOT NULL,
  `HDL` int(4) NOT NULL,
  `LDL` int(4) NOT NULL,
  `TRIGLI` int(4) NOT NULL,
  `TOTAL` int(4) NOT NULL,
  `CLASSE` int(1) NOT NULL,
  PRIMARY KEY (`ID`))
ENGINE=InnoDB DEFAULT CHARSET=latin1;
```

## Comando para enviar do Excel para base modelada no MYSQL:

- `load data local infile 'c:/exame.csv' into table exames fields terminated by ';' lines terminated by '\r\n' (`registro`, `sexo`, `idade`, `resultado`, `glicemia`, `hdl`, `ldl`, `triglic`, `total`, `classe`)`

## Comandos para classificação das redes:

- `update `exames` set classe=1`
- `update `exames` set ldl=total-(triglic/5)-hdl where ldl=0 and total>0 and triglic>0 and hdl>0 and triglic<=400`
- `update `exames` set classe=0 where idade>=20 and (glicemia between 50 and 126) and (total between 100 and 240) and (hdl between 15 and 100) and (ldl between 25 and 130) and (triglic between 10 and 150)`
- `update `exames` set classe=0 where idade<20 and (glicemia between 50 and 100) and (total between 60 and 170) and (hdl between 15 and 100) and (ldl between 25 and 110) and (triglic between 10 and 100)`

## Comandos para dados estatísticos:

- `select count(distinct(registro)) from `exames``
- `select count(distinct(registro)) from `exames` where sexo=1 and idade>19 and registro not in( select registro from `exames` where idade<20)`
- `select count(distinct(registro)) from `exames` where sexo=2 and idade>19 and registro not in( select registro from `exames` where idade<20)`
- `select count(distinct(registro)) from `exames` where idade<20`
- `select count(*) from `exames` where `classe`='1'`
- `select count(*) from `exames` where `classe`='0'`

## Comandos para exportar do MYSQL para arquivos de treinamento e teste:

- `select sexo,idade,glicemia,hdl,ldl,triglic,total,classe from `exames` where id <= 30003 into outfile 'c:/treina.csv' fields terminated by ';' lines terminated by '\n'`
- `select sexo,idade,glicemia,hdl,ldl,triglic,total,classe from `exames` where id > 30003 into outfile 'c:/teste.csv' fields terminated by ';' lines terminated by '\n'`

## APÊNDICE C – SCRIPTS EM MATLAB USADOS NO PROJETO

```
% -----
% Script: glicemia_lipidico.m
% EXAMES DE GLICEMIA E PERFIL LIPIDICO (HDL-C, LDL-C, TG, CT)
% Características: Sexo, Idade, Glicose em Jejum, HDL, LDL, Triglicérides,
%                  Colesterol Total
% Classe.....: Duas classes 0-LIBERA, 1-RETIDO
% Autor.....: Ademir Luiz do Prado
% -----
load 'C:\Users\ADEMIR\TREINA.txt' % carrega arquivo de TREINAMENTO
rede = mlp_tr(TREINA, [7 15 15])
RedeArquivoTREINA = mlp_ts(TREINA, rede)

load 'C:\Users\ADEMIR\TESTE.txt' % carrega arquivo de TESTE
RedeArquivoTESTE = mlp_ts(TESTE, rede)

save GlicemiaLipidico.dat

erroTREINA = erro_rede(TREINA, rede); % salva na variavel erroTREINA os
                                     % dados que a rede "CONFUNDIU"

erroTESTE  = erro_rede(TESTE, rede); % salva na variavel erroTESTE os dados
                                     % que a rede "CONFUNDIU"

save erroTREINA.xls erroTREINA -ascii; % salva os erros em arquivo excel

save erroTESTE.xls erroTESTE -ascii; % salva os erros em arquivo excel

% -----
% Script: erro_rede.m
% LISTA REGISTROS QUE A REDE NERONAL MLP DIVERGIU
% Recebe arquivo no primeiro parâmetro. No 2º parâmetro é a rede neuronal
% Autor.....: Ademir Luiz do Prado
% -----
function mret = erro_rede(argteste, rede);
    saida = mlp_ts(argteste, rede);
    m = length(argteste(1,:));
    id = find(saida.Result' ~= argteste(:,m));
    mret = argteste(id,:);

% -----
% Script: mlp_tr.m
% Treinamento da rede
% Adaptado de Roberto Tadeu Raittz
% -----
function mret = mlp_tr(w,epc)
m = length(w(1,:));
P = double(w(:,1:m-1))';
T = double(w(:,m))';
rand('seed', 672880951);
net = newff(P,T,epc); % create neuronal network
net.trainParam.epochs = 500;
%net.trainParam.show = 'off';
[net,tr] = train(net,P,T); % trains
mret = net;
```



```

% -----
% Script: mlp_ts.m
% Teste da rede
% Adaptado de Roberto Tadeu Raittz
% -----
function mret = mlp_ts(w,net)
m = length(w(1,:));
n = length(w(:,1));
P = double(w(:,1:m-1))';
T = double(w(:,m))';
nc = max(T);
testInputs = P;
testTargets = T;
out = round(sim(net,testInputs));
xout = out;% Get response of trained network
if min(T) == 1
    T = T-1;
    out = out - 1;
    nc = nc-1;
end
out(find(out>nc)) = nc;
out(find(out<0)) = 0;
conf = zeros(nc+1,nc+1);
for i=1:n
    conf(T(i)+1,out(i)+1) = conf(T(i)+1,out(i)+1)+1;
end
mret.conf = conf;
r = [];
for i=1:(nc+1)
    u = conf(i,i)/sum(conf(:,i));
    r = [r u];
end
certos = 100*(sum(out==T))/n;
mret.certos = certos;
mret.prop = r;
mret.Result = xout;

```

APÊNDICE D – EXEMPLO DO BANCO DE DADOS NO MYSQL

phpMyAdmin

Banco de Dados

glicemico\_lipidico (1)

glicemico\_lipidico (1)

exames

Servidor: localhost ▶ Banco de Dados: glicemico\_lipidico ▶ Tabela: exames

Visualizar

Estrutura

SQL

Procurar

Inserir

Exportar

Importar

Operações

Limpar

Eliminar

Mostrando registros 0 - 29 (60,006 total, Consulta levou 0.0006 segundos)

SELECT \* FROM "exames" LIMIT 0 , 30

☐ Profiling [ Editar ] [ Explicar SQL ] [ Criar código PHP ] [ Atualizar ]

no modo horizontal

registro(s) começando de 30

>>

Numero da página: 1

Ordenar pela chave: Nenhum

+ Opções

>

e repetindo cabeçalhos após 100 células

	ID	REGISTRO	SEXO	IDADE	RESULTADO	GLICEMIA	HDL	LDL	TRIGLI	TOTAL	CLASSE	
<input type="checkbox"/>	<input type="checkbox"/>	1	590	1	53	2006-09-19	109	69	98	107	188	0
<input type="checkbox"/>	<input type="checkbox"/>	2	590	1	54	2007-04-26	81	85	76	83	178	0
<input type="checkbox"/>	<input type="checkbox"/>	3	590	1	55	2008-04-22	122	59	160	188	257	1
<input type="checkbox"/>	<input type="checkbox"/>	4	590	1	55	2008-10-16	123	64	196	152	290	1
<input type="checkbox"/>	<input type="checkbox"/>	5	590	1	56	2009-06-03	115	74	102	138	204	0
<input type="checkbox"/>	<input type="checkbox"/>	6	2798	1	58	2010-07-30	93	59	138	96	216	1
<input type="checkbox"/>	<input type="checkbox"/>	7	4138	1	53	2006-11-20	276	114	335	163	482	1
<input type="checkbox"/>	<input type="checkbox"/>	8	4464	1	59	2006-09-06	82	53	97	97	169	0
<input type="checkbox"/>	<input type="checkbox"/>	9	4464	1	59	2006-09-16	100	49	88	83	154	0
<input type="checkbox"/>	<input type="checkbox"/>	10	4871	1	59	2007-02-02	102	65	88	76	168	0
<input type="checkbox"/>	<input type="checkbox"/>	11	4871	1	59	2007-06-18	104	59	76	59	147	0
<input type="checkbox"/>	<input type="checkbox"/>	12	5304	1	49	2007-05-23	99	54	176	87	247	1
<input type="checkbox"/>	<input type="checkbox"/>	13	5304	1	50	2008-03-31	99	43	169	128	238	1
<input type="checkbox"/>	<input type="checkbox"/>	14	5410	1	48	2006-09-21	87	55	133	165	221	1
<input type="checkbox"/>	<input type="checkbox"/>	15	5410	1	49	2007-10-18	65	37	66	67	116	0
<input type="checkbox"/>	<input type="checkbox"/>	16	5410	1	50	2008-03-20	57	34	62	83	113	0
<input type="checkbox"/>	<input type="checkbox"/>	17	5410	1	50	2008-08-12	89	47	149	141	224	1
<input type="checkbox"/>	<input type="checkbox"/>	18	5410	1	50	2008-10-01	81	39	145	132	210	1
<input type="checkbox"/>	<input type="checkbox"/>	19	6025	1	52	2009-02-26	106	43	144	169	221	1
<input type="checkbox"/>	<input type="checkbox"/>	20	6025	1	53	2010-03-05	107	56	129	109	207	0
<input type="checkbox"/>	<input type="checkbox"/>	21	6084	1	55	2010-01-14	99	48	142	120	214	1
<input type="checkbox"/>	<input type="checkbox"/>	22	6718	1	65	2007-10-19	104	39	188	167	260	1

## APÊNDICE E – EXEMPLO DA INTERFACE GRÁFICA: REDE NEURONAL MLP

The screenshot displays the MATLAB 7.10.0 (R2010a) environment. The Command Window shows the execution of a script named 'rede.m', which performs the following steps:

```

performParam: (none)
trainParam: .show, .showWindow, .showCommandLine, .epochs,
            .time, .goal, .max_fail, .mem_reduc,
            .min_grad, .mu, .mu_dec, .mu_inc,
            .mu_max

weight and bias values:

IW: {4x1 cell} containing 1 input weight matrix
LW: {4x4 cell} containing 3 layer weight matrices
b: {4x1 cell} containing 4 bias vectors

other:
    name: ''
    userdata: (user information)

RedeArquivoTREINA =
    conf: [2x2 double]
    certos: 99.6200
    prop: [0.9940 0.9976]
    Result: [1x30003 double]

RedeArquivoTESTE =
    conf: [2x2 double]
    certos: 98.7435
    prop: [0.9791 0.9925]
    Result: [1x30003 double]
  
```

The Workspace window shows the following variables:

Name	Value	Min
RedeArquivoTESTE	<1x1 struct>	
RedeArquivoTREI...	<1x1 struct>	
TESTE30	<30003x8 double>	0
TREINA30	<30003x8 double>	0
erroTESTE	<379x8 double>	0
erroTREINA	<116x8 double>	0
rede	<1x1 network>	

The Command History window shows the command 'glicemia\_lipidico' entered at 31/01/11 06:57.

## APÊNDICE F – EXEMPLO DA INTERFACE GRÁFICA: REDE NEURONAL FAN

EasyFAN

Arquivo Configurações Ajuda

**Opções**

Adicionar Arquivo

Limpar Conjunto

Salvar Conjunto

Configurar Parâmetros

**Opções de Rede**

Validar Rede

Matriz de Confusão

Visualizar Conjunto de Validação

Matriz de Confusão: Teste

0 1

0 10940 232

1 2050 16781

Quantidade de Ocorrências

Porcentagem

Média Harmônica da Época: 93.51853942871094

Máximo do Mínimo da Época: 89.11369323730469

Média Aritmética da Época: 92.39409637451172

Sair

Conjunto de Validação		Nº de Padrões:		Nº de Classes:								
Característica 1	Característica 2	Característica 3	Característica 4	Característica 5	Característica 6	Característica 7	Cl.					
1	44	91	63	114	73	192	0					
1	52	103	53	258	123	336	1					
1	59	97	42	101	68	157	0					
1	30	96	43	117	88	178	0					
1	30	224	35	115	136	177	1					
1	71	113	39	141	128	206	1					
1	71	110	38	116	140	182	0					
1	72						1					
1	32						0					
1	56						0					
1	52						0					
1	56						1					
2	52						0					
1	44						0					
1	56						1					
1	57						1					
1	57						1					
1	80						1					
1	81						1					
1	47						0					
1	47						0					

Visualizar Conjunto de Validação

Próxima Página

Anterior

Página: 1 / 101

Época Validada: --

Tipo de Rede Usada: Melhor Média Harmônica

Média harmônica: 093.5185

Máximo do mínimo: 089.1137

Média aritmética: 092.3941

Acertos: 27721

Erros: 2282

EasyFAN Treinamento Teste Validação Classificação

## ANEXO – APROVAÇÃO DO PROJETO NO COMITÊ DE ÉTICA



Curitiba, 30 de setembro de 2010.

Ilmo (a) Sr. (a)  
**Ademir Luiz do Prado**  
Neste

Prezado Pesquisador:

Comunicamos que o Projeto de Pesquisa intitulado “ESTUDO E APLICAÇÃO DE REDES NEURONAIS ARTIFICIAIS PARA APOIO A DECISÃO NA LIBERAÇÃO DE EXAMES DE COLESTEROL”, foi analisado e aprovado pelo Comitê de Ética em Pesquisa em Seres Humanos, em reunião realizada no dia 28 de setembro de 2010. O referido projeto atende aos aspectos das Resoluções CNS 196/96, e demais, sobre Diretrizes e Normas Regulamentadoras de Pesquisa Envolvendo Seres Humanos do Ministério da Saúde.

**CAAE: 0253.0.208.000-10**  
**Registro CEP: 2333.227/2010-09**

Conforme a Resolução 196/96, solicitamos que sejam apresentados a este CEP, relatórios sobre o andamento da pesquisa, bem como informações relativas às modificações do protocolo, cancelamento, encerramento e destino dos conhecimentos obtidos.

Data para entrega do primeiro relatório: 30 de março de 2011.

Atenciosamente,

A handwritten signature in black ink, appearing to read 'Renato Tambara Filho'.

**Renato Tambara Filho**

Coordenador do Comitê de Ética em Pesquisa  
em Seres Humanos do Hospital de Clínicas/UFPR